INVESTIGACION Y CIENCIA

Edición en español de **SCIENTIFIC AMERICAN**



CORAL

Octubre 1979

200 Ptas.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 37
Octubre 1979

INVESTIGACION Y CIENCIA

Edición en español de SCIENTIFIC AMERICAN

INVESTIGACION Y CIENCIA

DIRECTOR Francisco Gracia Guillén
REDACCION José María Valderas Gallardo (Redactor Jefe)
Encarna de Blas (Secretaria de Redacción)
PRODUCCION César Redondo Zayas
PROMOCION Y PUBLICIDAD Elena Sanchez-Fabrés
PROMOCION EXTERIOR Pedro Clotas Cierco

EDITA Prensa Científica, S.A.
Calabria, 235-239
Barcelona-29 (ESPAÑA)

Colaboradores de este número:

Asesoramiento y traducción:

Carlos Oppenheimer y Manuel Blanco: *Tests bacterianos de substancias potencialmente cancerígenas;* Joaquín García y Javier Viñas: *Progresos hacia un reactor de fusión tokamak;* Antonio Blanco: *Una aldea de colonos preneolíticos a orillas del Eúfrates;* Joandomènec Ros: *Corales y arrecifes coralinos;* Manuel Puigcerver: *Supernovas y formación de estrellas;* Ramón Pascual: *Problemas físicos con muchas escalas de longitud;* Joandomènec Ros: *El control hormonal del comportamiento en un lagarto;* Luis Bou: *Juegos matemáticos;* José E. Myro y Alicia Menéndez: *Taller y laboratorio.*

Libros:

Carlos M. Cuadras, Manuel de la Morena y Antonio Escarré



LA PORTADA

La fotografía de la portada es un primer plano de una colonia viva del coral cerebriforme *Diploria*, situado en el Elkhorn Reef, en la Small Hope Bay, en aguas de la isla Andros, de las Bahamas. Las fisuras del esqueleto madrepórico están tapizadas con pólipos vivos, que se retraen hacia el interior de los surcos durante el día, como se aprecia en la fotografía. Al amanecer y al atardecer los pólipos emergen para alimentarse, inmovilizando a crustáceos diminutos y a otros miembros del zooplancton con las células urticantes de sus tentáculos. La notable capacidad que tienen los pólipos vivos para precipitar enormes esqueletos de carbonato cálcico, mediante los cuales construyen estructuras robustas y resistentes al oleaje, depende de su íntima asociación con unas algas simbiontes llamadas zooxantelas (véase "Corales y arrecifes coralinos", de Thomas F. Goreau, Nora I. Goreau y Thomas J. Goreau, en este número). A pesar de la reducida fertilidad de los mares tropicales, los arrecifes de coral albergan una enorme diversidad de especies vegetales y animales, y constituyen, probablemente, el ecosistema más complejo.

Suscripciones:
Prensa Científica, S.A.
Calabria 235-239
Barcelona-29 (España)
Teléfono 322 05 51 ext. 41

Condiciones de suscripción:
España:
Un año (12 números): 2200 pesetas
Extranjero:
Un año (12 números): 36 U.S.$
Ejemplar atrasado ordinario:
225 pesetas
Ejemplar atrasado extraordinario:
340 pesetas

Distribución para España:
Distribuciones de Enlace, S.A.
Ausias March, 49; Barcelona-10

Distribución para los restantes países:
Editorial Labor, S.A.
Calabria, 235-239; Barcelona-29

Publicidad:
Madrid:
Gustavo Martínez Ovin
Avda. de Moratalaz, 137; Madrid-30
Tel. 430 84 81
Cataluña:
Miguel Munill
Balmes, 191, 2.°, 3.ª; Barcelona-6
Tels. 218 44 45 y 218 40 86

Controlado por O.J.D.



PROCEDENCIA DE LAS ILUSTRACIONES

Fotografía de la portada de Douglas Faulkner

Página	Fuente
7	Raymond Devoret, Laboratorio de Enzimologia, C.N.R.S., Gif-sur-Yvette
8	Ilil Arbel
9	Dirk Bootsma, Universidad Erasmo de Rotterdam (*arriba*); Ilil Arbel (*abajo*)
10-11	Ilil Arbel
12	Raymond Devoret, Laboratorio de Enzimologia, C.N.R.S., Gif-sur-Yvette
13-15	Ilil Arbel
18-22	George V. Kelvin
23	Ned R. Sauthoff
24-25	George V. Kelvin
26	George V. Kelvin (*arriba*), Ned R. Sauthoff (*abajo*)
27-29	George V. Kelvin
33	Andrew M. T. Moore
34	Andrew Tomko
35	Andrew M. T. Moore
36-37	Alan D. Iselin
38-39	Andrew Tomko
40	Alan D. Iselin
49	Douglas Faulkner
50	Peter Arnold, Inc.
51	Thomas F. Goreau
52-53	Tom Prentiss
54	Tom Prentiss (*arriba y abajo, derecha*) Robert K. Trench (*abajo, izquierda*)
55-57	Tom Prentiss
58	Thomas F. Goreau
63	William Herbst, Wesleyan University
64-66	Walken Graphics
70	Huberto Gerola y Philip E. Seiden, Thomas J. Watson Research Center, International Business Machines Corporation
73	LCR Graphics, Inc.
74-78	Gabor Kiss
80-81	LCR Graphics, Inc.
82-92	Gabor Kiss
95	Fritz Goro
96-98	Patricia J. Wynne
100	Joan I. Morrell, Rockefeller University
102-103	Patricia J. Wynne
105-112	M. Puigcerver y Toni Sellés
113	M. Puigcerver
115-116	M. Puigcerver y Toni Sellés
117	M. Puigcerver y Toni Sellés (*arriba*); M.Puigcerver (*abajo*)
118	M. Puigcerver y Toni Sellés
122-125	Ilil Arbel
128-129	Ben Rose
130-132	Michael Goodman
133	Michael Goodman (*arriba*), Ben Rose (*abajo*)

ISSN 0210-136X
Dep. legal: B. 38.999-76
Fotocomposición: Tecfa
Guipúzcoa, 36 (local 1) Barcelona-20
Cayfosa. Santa Perpetua de Moguda
Barcelona
Printed in Spain - Impreso en España

Los autores

RAYMOND DEVORET ("Tests bacterianos de substancias potencialmente cancerígenas") pertenece al equipo investigador del Laboratorio de Enzimología en el Centre National de la Recherche Scientifique (CNRS) en Gif-sur-Yvette, Francia. Tras graduarse en la Facultad de medicina de la Universidad de París, trabajó como médico especializado en la protección de los riesgos derivados del trabajo con radiaciones. Simultáneamente, se graduó en física en la Universidad de París y, en el Instituto del Radio, estudió los efectos de los rayos alfa sobre la bacteria *Escherichia coli*. Desde 1961, fecha en que pasó a pertenecer al CNRS, las investigaciones de Devoret han estado dedicadas a elucidar los mecanismos de la inducción de virus latentes en bacterias y de la carcinogénesis química.

HAROLD P. FURTH ("Progresos hacia un reactor de fusión tokamak") es profesor de ciencias astrofísicas en la Universidad de Princeton y director asociado en el Laboratorio de Física de Plasmas en Princeton. Trabajó como graduado en física del estado sólido y en física de altas energías en la Universidad de Harvard, recibiendo su grado de doctor en 1960. Después se unió al grupo investigador sobre la fusión controlada en el laboratorio Lawrence Livermore, llegando a ser el director de un grupo que estudiaba el confinamiento toroidal magnético, una aproximación para un reactor de fusión. En 1967 se trasladó a Princeton, donde ha iniciado un número de importantes experimentos sobre confinamiento toroidal y realizado una dilatada investigación sobre la teoría de las inestabilidades del plasma.

ANDREW M. T. MOORE ("Una aldea de colonos preneolíticos a orillas del Eúfrates") enseña prehistoria del Viejo Mundo en la Universidad Arizona. Cursó historia moderna en la Universidad de Oxford, y después se formó en arqueología en el Instituto de Arqueología de la Universidad de Londres. En 1969 obtuvo una beca en la Escuela Británica de Arqueología de Jerusalén, y durante los dos años siguientes vivió en el Próximo Oriente, estudiando colecciones de museos y visitando yacmientos antiguos. En 1973, se le concedió una plaza de *fellow* de la fundación de Wainwright en Oxford. Obtuvo el doctorado el pasado año. Moore ha realizado muchos trabajos de campo en el Próximo Oriente; participó en excavaciones en Jerusalén, dirigidas por la finada Dame Kathleen Kenyon, en 1966; y en las de Knossos, en Creta, por J. D. Evans, en 1969. De 1972 a 1973 Moore estuvo al frente de la importante excavación de Tell Abu Hureyra en Siria.

THOMAS F. GOREAU, NORA I. GOREAU y THOMAS J. GOREAU ("Corales y arrecifes coralinos") han sido pioneros en la investigación de la biología y de la ecología de los arrecifes de coral en Jamaica. Thomas F. Goreau murió en 1970. Hijo de los fotógrafos Fritz y Grete Goro, nació en Alemania. Vivió en Francia y los Estados Unidos. Después de graduarse en la Universidad de Clark realizó trabajos de investigación en ecología en la Universidad de Yale. Goreau se interesó por primera vez por la ecología de los arrecifes de coral en 1947, aunque la primera oportunidad para proseguir esta línea de investigación no surgió hasta 1951. Al recibir su doctorado en 1956 instituyó un proyecto de investigación a largo plazo de los arrecifes coralinos jamaicanos, que continuó hasta 1967. En marzo de 1970, abrió un nuevo laboratorio marino en Discovery Bay. La viuda de Goreau, Nora Goreau, es en la actualidad jefe de investigación en biología marina en la Universidad de las Antillas. Nacida en Panamá, se educó en la Law School de la Universidad de Panamá, en la Universidad de Iowa y en la Universidad de Paul. Posteriormnte realizó investigaciones en neurofisiología en la Universidad de Chicago. Thomas J. Goreau, su hijo mayor, está haciendo su doctorado en geología en la Universidad de Harvard. Estudió física y astronomía en el Instituto de Tecnología de Massachusettss y biología y oceanografía en Yale, la Woods Hole Oceanographic Institution y Harvard.

WILLIAM HERBST y GEORGE E. ASSOUSA ("Supernovas y formación de estrellas") han colaborado en estudios de formación de estrellas con telescopios ópticos y radiotelescopios. Herbst es profesor adjunto de astronomía en la Wesleyan University. Efectuó sus estudios de primer ciclo universitario en astrofísica en la Universidad de Princeton y obtuvo su doctorado en astronomía en la Universidad de Toronto en 1974. Tras obtener una beca postdoctoral en la Universidad de Nueva York, fue durante dos años miembro de la Institución Carnegie de Washington. Herbst ingresó en el cuerpo docente de la Wesleyan en 1978. Assousa es profesor de investigación en astrofísica en el Departamento de Magnetismo Terrestre de la Institución Carnegie. Es árabe palestino y nació en Jerusalén en 1936, asistiendo en Ramallah, Jordania, a la American Friends School. Después marchó a los Estados Unidos para completar su formación, obteniendo su grado de Bachelor en el Earlham College, su grado de Master en la Universidad de Columbia y su doctorado en física nuclear experimental en la Universdad de Florida en 1968. Es miembro de la Fundación para la Reconciliación Arabe-Israelí.

KENNETH G. WILSON ("Problemas de la física con muchas escalas de longitud") es profesor James A. Weeks de ciencia física en la Universidad de Cornell. Se graduó en la de Harvard y se doctoró en el Instituto de Tecnología de California en 1961. La teoría del grupo de renormalización, su contribución principal a la física, surgió de su intento por comprender la teoría cuántica de campos.

DAVID CREWS ("El control hormonal del comportamiento en un lagarto") es profesor asociado de biología y psicología en la Universidad de Harvard y colaborador del Museum of Comparative Zoology. Realizó sus estudios en la Universidad de Maryland y obtuvo su título de doctor en comportamiento animal por la Universidad de Rutgers. Posteriormente realizó investigaciones postdoctorales en la Universidad de California en Berkeley y en Harvard, donde en 1976 se incorporó a la facultad.

MANUEL PUIGCERVER ("Atmósfera y contaminación atmosférica") se licenció en Ciencias en la Universidad de Valencia en 1946, ingresando inmediatamente en el Servicio Meteorológico. En 1957 ganó una beca de la Administración de Cooperación Internacional de los Estados Unidos, obteniendo el Master of Science en Meteorología en la Universidad de Chicago. Se doctoró en la Universidad de Barcelona en 1962. Durante todo el año 1961 estuvo en una base antártica desempeñando las misiones de director científico y meteorólogo. Desde 1967 es catedrático de Física de la Atmósfera en la Universidad de Barcelona. Ha trabajado en electricidad atmosférica, meteorología antártica y mesometeorología. Su interés se centra hoy en los intercambios energéticos y de materia entre superficie y atmósfera incluyendo la contaminación atmosférica.

Hace...

José M.ª López Piñero

...cuatrocientos años

Pedro Sarmiento de Gamboa inició su expedición al Estrecho de Magallanes con la finalidad de reconocer la zona, fortificar sus orillas y fundar un asentamiento. Aunque no pudo llevar a cabo estos dos últimos objetivos, escribió memoriales y una relación y derrotero del viaje de notable valor geográfico.

Tras formarse en matemática y arte de navegar en la metrópoli, Sarmiento se había trasladado a América dos décadas antes. El virreinato del Perú le encargó que escribiera una historia de los incas, con motivo de lo cual entró en relación con personalidades como Tupac Yupanqui quien, entre otras cosas, le habló de las noticias que los incas tenían acerca de las tierras del sur del Pacífico. En 1568 fue nombrado piloto de uno de los dos barcos de una expedición, dirigida por Alvaro de Mendaña, que se proponía precisamente explorar dicha parte del mundo. La travesía estuvo llena de incidentes pero, gracias a la pericia de Sarmiento, encargado de fijar el derrotero, los expedicionarios pudieron regresar felizmente. Entre los resultados conseguidos, el más importante fue el descubrimiento de las islas Salomón, cuya primera descripción geográfica se debe a Sarmiento.

Después de sus viajes al Estrecho de Magallanes, Sarmiento volvió a España. En Sevilla, se dedicó al trazado de cartas náuticas y al diseño de instrumentos para la navegación, colaborando con Rodrigo Zamorano, que entonces era Piloto Mayor y catedrático de Cosmografía en la Casa de la Contratación. Presentó también un informe sobre la corrección de la longitud de la isla Asunción y otro acerca de un instrumento astronómico de su invención. Como cartógrafo y diseñador de instrumentos científicos fue muy apreciado en toda Europa.

...trescientos años

Publicó sus dos obras Juan Bautista Orivai de Monreal, una de las figuras más destacadas de la mentalidad tradicional, intransigente a las novedades, de la medicina española durante las décadas centrales del siglo XVII. Ocupó varias cátedras médicas en la Universidad de Valencia, entre ellas la de Hipócrates, así como la de lengua griega. La escuela médica valenciana, que había sido durante el Renacimiento un importante núcleo de renovación científica, era entonces un reducto académico del galenismo reaccionario, similar al de la Sorbona en Paris. Matías García, que ocupaba la cátedra de anatomía, se dedicó a atacar las novedades anatómicas y fisiológicas, en especial la doctrina de la circulación de la sangre. Orivai hizo otro tanto con cualquier disidencia de la patología y la terapéutica tradicionales.

A esta postura responde exactamente la primera de sus obras, titulada *Propugnaculum Hippocraticae ac Galenicae doctrinae de febrium putridarum,* que fue impresa en Lyon. Dividida en tres partes, la primera es un debate escolástico sobre el uso de los purgantes en ese tipo de fiebres; la segunda, un ataque al médico catalán Bernardo Caxanes, que había criticado la exagerada inclinación de los valencianos a la práctica de la sangría; la tercera es otro escrito polémico, asimismo relacionado con la sangría, dirigido contra el sevillano Gaspar Caldera de Heredia, importante galenista abierto parcialmente a las novedades. A lo largo del libro se polemiza también con otras figuras de mentalidad moderada, como Gaspar Bravo de Sobremonte y Pedro Miguel de Heredia, catedráticos en Valladolid y Alcalá, respectivamente.

La segunda obra, por el contrario, versa sobre un problema de medicina práctica. Impresa en Zaragoza, se titula *Teatro de la verdad y claro manifiesto del conocimiento de las enfermedades de la ciudad de Orihuela del año de 1678.* En ella demuestra que no se trataba de peste ni de otro contagio, sino de "calenturas malignas con forma vulgar, nacidas de pobreza, hambre, malos alimentos y destemplanza del tiempo". Frente a los esquemas históricos simplistas, es una prueba de que los galenistas intransigentes podían ser excelentes observadores clínicos y epidemiológicos.

...doscientos años

Juan Andrés Morell publicó, en italiano y en castellano, su *Lettera... sopra una Dimostrazione del Galileo.* Era el segundo de los estudios que dedicaba a esta figura clave del nacimiento de la ciencia moderna, ya que tres años antes había aparecido su *Saggio della Filosofia del Galileo.* Juan Andrés era un jesuita alicantino, que había sido profesor en los colegios de Valencia y Gandía hasta la expulsión de la Compañía en 1767. Marchó entonces a Italia, donde permaneció hasta su muerte en 1817, dedicado a la enseñanza y a la redacción de obras filosóficas, históricas y científicas. Una de las más tempranas fue la resolución de un problema de hidráulica propuesto por la Academia de Mantua en 1774. Mayor interés tienen sus estudios sobre Galileo, que se encuentran entre los primeros acerca del tema publicados desde una ideología católica. Víctor Navarro ha destacado que contrapone la figura de Galileo a las de Descartes y Francis Bacon, considerándolo como el verdadero fundador de la ciencia moderna. También ha recordado el juicio de Clelia Pigheti, que encuentra sospechosa la defensa de Andrés en un momento en el que había alcanzado plena vigencia la física newtoniana, que incorporaba por completo la obra del pisano. En cualquier caso, Andrés disimuló las implicaciones teológicas del proceso de Galileo, achacando la persecución a la "oscuridad de la época".

Posteriormente, Juan Andrés publicó una amplia historia de la cultura en siete volúmenes, titulada *Dell'origine, progressi e stato attuale d'ogni letteratura* (1782-99), en la que dedicó atención preferente a la evolución de las matemáticas y las ciencias de la naturaleza. La traducción castellana, que hizo su hermano Carlos Andrés, apareció en Madrid entre 1784 y 1806, siendo declarada libro de texto en los Reales Estudios.

...cien años

Nació en Madrid el zoólogo Angel Cabrera Latorre. Su padre, Juan Bautista Cabrera, fue el primer obispo de la Iglesia Episcopal Española. Estudió filosofía y, más tarde, ciencias naturales en la Universidad de Madrid. Su interés por la biología fue muy temprano, ya que ingresó a los diecisiete años en la Sociedad Española de Historia Natural y colaboró poco después con Marcos Jiménez de la Espada en el estudio de los mamíferos americanos traídos por la Expedición Científica al Pacífico de 1862-65.

A los mamíferos dedicó asimismo la mayor parte de su obra científica, que se desarrolló en el Museo de Ciencias Na-

turales de Madrid, entre 1902 y 1925, y en el Departamento de Paleontología del Museo de La Plata, a partir de la última fecha citada.

Entre sus primeros trabajos, destaca *El concepto de tipo en Zoología y los tipos de mamíferos del Museo de Ciencias Naturales* (1912) y un *Catálogo metódico* (1912-13) de las colecciones de mamíferos del mismo centro. Más tarde, publicó numerosos artículos en revistas especializadas de todo el mundo y fue elegido miembro de prestigiosas instituciones científicas, como la Zoological Society de Londres y la Biological Society de Washington.

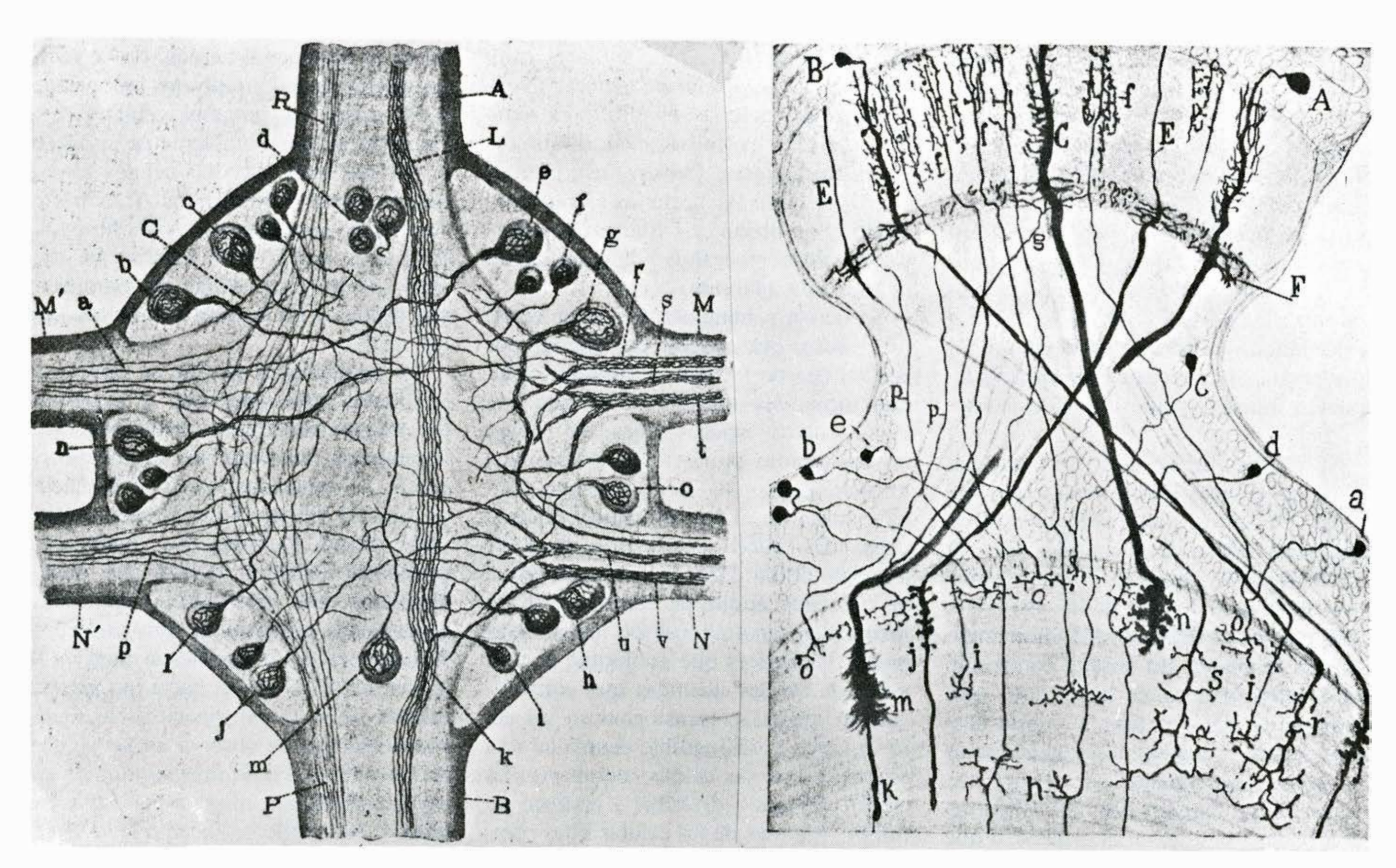
Angel Cabrera dedicó, por otro lado, notable atención a la divulgación científica. Aparte de una amplia serie de artículos en revistas como *Alrededor del mundo*, su principal contribución en este campo fue una *Historia natural. Vida de los animales, de las plantas y de la tierra* (1927) que alcanzó una amplia y merecida popularidad.

...cincuenta años

Domingo Sánchez Sánchez inició, con sus investigaciones sobre el ganglio cerebroide de la *Blata orientalis* o cucaracha, la última fase de su obra en torno a la neurohistología de los invertebrados.

Nacido en la localidad salmantina de Fuenteguinaldo, había estudiado ciencias naturales en la Universidad de Madrid, obteniendo la licenciatura en 1885. El mismo año marchó a Filipinas con un cargo de "colector zoológico". Además de recoger materiales de la fauna local, publicó una monografía sobre un parásito de los cafetales filipinos y preparó su tesis doctoral, que versó sobre casi un centenar de mamíferos de las islas. De regreso a España en 1898, terminó en la Universidad de Madrid los estudios de medicina que había iniciado en Manila. Con este motivo, entró en relación con Santiago Ramón y Cajal, hecho que condicionó el resto de su biografía científica. Antes de acabar la nueva licenciatura, se incorporó al laboratorio dirigido por Cajal y, desde el principio, se centró en la investigación histológica de los invertebrados y, a partir de 1904, en el estudio de su sistema nervioso. En este último año comenzó a utilizar el método del nitrato de plata reducido, con el que consiguió una descripción exhaustiva de la estructura de la cadena ganglionar y de los nervios periféricos de los hirudíneos. Sus hallazgos contribuyeron a consolidar la teoría neuronal frente a la reticularista, apoyada principalmente en datos procedentes de los invertebrados.

La obra de Domingo Sánchez culminó con sus trabajos sobre neurohistología de los insectos. Sobresalen, a este respecto, un volumen consagrado a la retina y los centros ópticos de los insectos superiores (1915), que firmó en colaboración con Cajal, y los estudios de la última fase de su vida, a la que hemos aludido, centrados en la *Blata orientalis* y en la *Apis mellifica* o abeja obrera.



Estructuras neurohistológicas de los invertebrados, según Domingo Sánchez. **A la izquierda**, *corte horizontal de un ganglio medio del* Hirudo medicinalis. A, *cordón longitudinal anterior;* B, *ídem posterior;* C, *cápsula interna;* M, M', *nervios anteriores con las fibras independientes* (a) *y los manojos* (r, s, t) *que por ellos pasan;* N, N', *nervios posteriores con las fibras independientes* (p) *y los manojos* (u, v) *que por ellos pasan;* b, c, *células del foco antero-lateral izquierdo;* d, *ídem del central anterior;* e, f, g, *ídem del antero-lateral derecho;* n, o, *ídem de los laterales medios o inter-radiculares;* h, i, *ídem del postero-lateral derecho;* j, l, *ídem del íd. izquierdo;* m, *ídem del central posterior;* L, *manojo central longitudinal derecho que se encorva anterior y posteriormente* (k) *para formar los focos celulares centrales respectivos;* R, P, *fibras cordonales anteriores y posteriores.* **A la derecha,** *algunos elementos de la retina de la abeja. Sección horizontal.* A, C, *monopolares colosales terminadas mediante pie dentellado y grosero* (n); B, *monopolares acabadas a favor de ramaje difuso;* E, *monopolares continuadas con fibras descendentes* (K, J); a, b, *células en* T; e, *célula en* T, *cuyo penacho descendente acaba en plano más profundo que los elementos congéneres;* h, *fibras centrífugas largas terminadas en la retina intermedia*

Tests bacterianos de substancias potencialmente cancerígenas

Un conjunto de nuevos ensayos de corta duración puede identificar agentes que causan lesiones en el ADN, hecho inicial de la carcinogénesis química. Los tests son de gran valor para clarificar el daño ambiental en el ADN

Raymond Devoret

El hecho de que ciertos agentes físicos y químicos presentes en el medio ambiente incrementen la incidencia del cáncer ha cobrado gran interés. Es inmediato que una adecuada limitación de la exposición del hombre a substancias cancerígenas salvaría numerosas vidas. Se hace, pues, urgente identificar qué substancias del medio ambiente son las potencialmente cancerígenas. En lo que se refiere a los carcinógenos químicos no es una labor fácil. Se calcula que más de 50.000 productos químicos distintos elaborados por el hombre son de uso comercial e industrial corriente; cada año irrumpen en el mercado entre 500 y 1000 productos químicos nuevos. Los ensayos habituales efectuados con animales para probar una potencial capacidad cancerígena requieren mucho tiempo y grandes sumas de dinero.

Afortunadamente, hay una alternativa a los clásicos experimentos con animales. Podemos sacar provecho de la profunda unidad de la materia viva recurriendo a las bacterias como organismos de ensayo. En vez de los dos o tres años que se necesitan en los experimentos con animales, un ensayo bacteriano sobre capacidad cancerígena dura sólo unas pocas horas o días, y cuesta bastante menos. Un eficaz test con bacterias ha sido desarrollado por Bruce N. Ames, de la Universidad de California en Berkeley, basado en la capacidad que ciertos productos químicos tienen de provocar mutaciones en las bacterias. Más recientemente, mis colegas del Centre National de la Recherche Scientifique en Gif-sur-Yvette y yo hemos ideado un grupo de tests basados en la capacidad de ciertos productos químicos de inducir la proliferación, en bacterias, de un virus aletargado (en estado latente). El ensayo de Ames y los nuestros, además de proporcionar un medio de identificación de productos químicos peligrosos, constituyen nuevos y poderosos instrumentos para el conocimiento y la comprensión de los puntos fundamentales del proceso carcinogénico iniciado por productos químicos.

El cáncer es una enfermedad propia de organismos pluricelulares muy evolucionados, como el hombre, mientras que las bacterias, en el otro extremo de la escala evolutiva, son diminutos seres unicelulares. Pudiera, así, parecer paradójico que las bacterias sirvan para identificar substancias que provocan el cáncer. Pero, en realidad, la paradoja no es más que aparente.

Se tiende a imaginar el cáncer como un tumor que puede extenderse por todo el cuerpo formando múltiples colonias tumorales (metástasis). Es esta una visión clínica, macroscópica, del cáncer en un estudio pluricelular, puesto que un solo gramo de tumor maligno contiene hasta un millón de células cancerígenas. El cáncer comienza a nivel de una sola célula. Una determinada célula de un tejido adulto se desarrolla de tal forma que, apartándose de las estrictas reglas fisiológicas que gobiernan el conjunto de células idénticas que constituyen un tejido, se transforma en un defecto único y distinguible dentro de una estructura por lo demás monótona. La célula empieza a dividirse y aparece un tumor. Algunas de las células hijas rompen la barrera del tejido, invaden los tejidos adyacentes y normalmente forman metástasis en puntos distantes. Las células cancerosas tienen una gran ventaja selectiva ya que escapan al destino programado para la mayoría de las células normales: envejecer y morir. Para aquellas, el cuerpo entero es un medio de cultivo en el que germinan, muriendo finalmente con el cuerpo que matan.

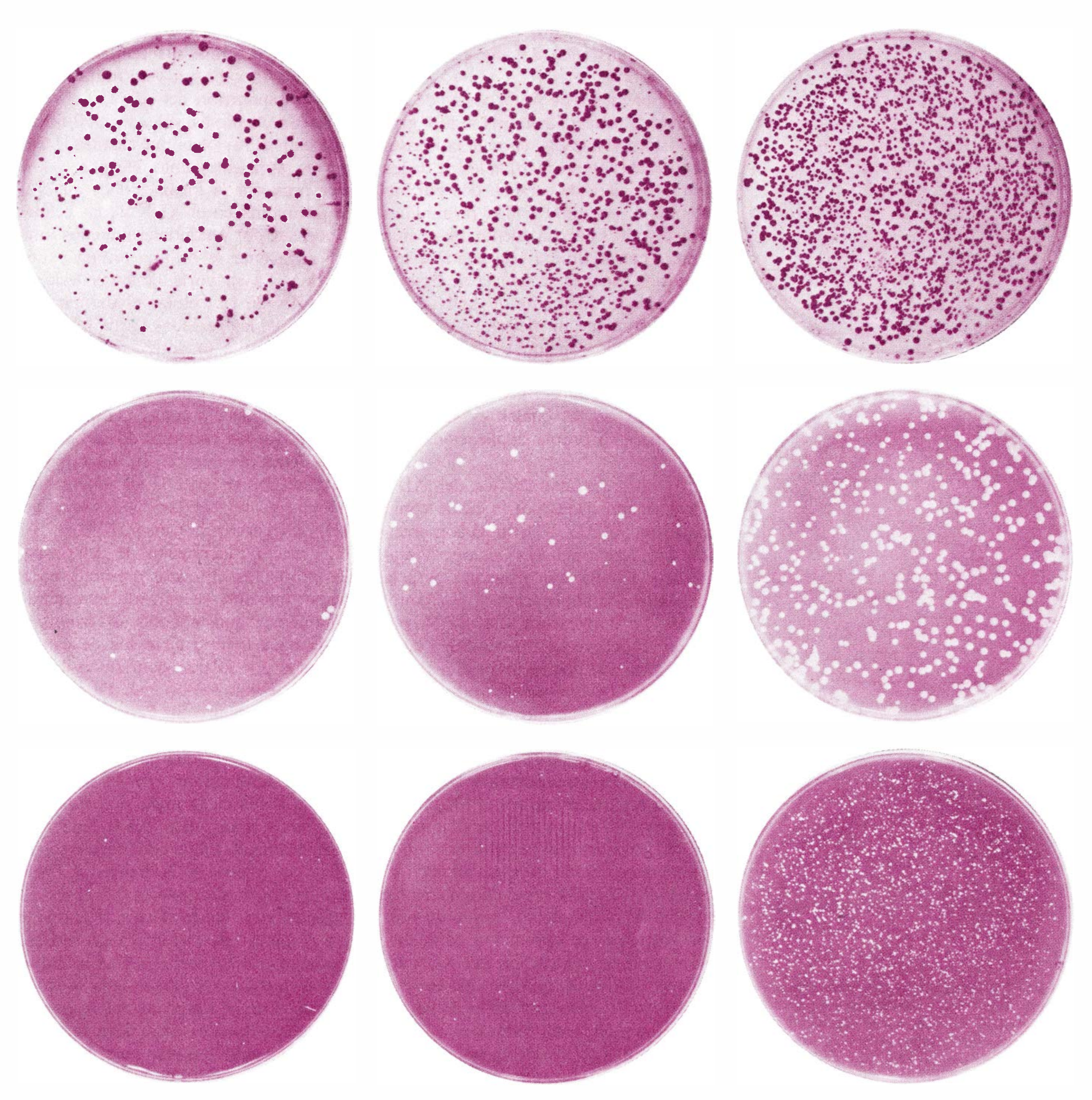
Ciertos agentes físicos y químicos del medio ambiente provocan el cáncer dañando el ADN, el material hereditario de la célula. La lesión del ADN inicia un complejo proceso celular que, en las células de los mamíferos, puede conducir, con el tiempo, a un estado canceroso. Los agentes que lesionan el ADN son, por tanto, cancerígenos en potencia. El ADN es el material hereditario de todas las células vivas, y tanto las lesiones del ADN como los procesos celulares que las reparan son notablemente similares en las bacterias y en las células humanas; lo que es perjudicial al ADN bacteriano daña igualmente el ADN humano. Esta es la justificación teórica que permite sustituir las células de mamíferos por bacterias en la detección de lesiones del ADN.

La justificación teórica se apoya en resultados experimentales y prácticos: los ensayos bacterianos distinguen entre substancias cancerígenas y no-cancerígenas, ya conocidas, con una fiabilidad de más del 90 por ciento, y han identificado como carcinógenos potenciales productos químicos nuevos que, posteriormente, han mostrado su naturaleza cancerígena en pruebas con animales. Naturalmente, las manifestaciones de la lesión del ADN en las bacterias difieren mucho de la transformación en estado canceroso que se observa en las células de mamíferos. En compensación, los ensayos bacterianos, al ser mucho más rápidos y económicos, hacen que la detección de carcinógenos a gran escala pueda considerarse ya un objetivo alcanzable.

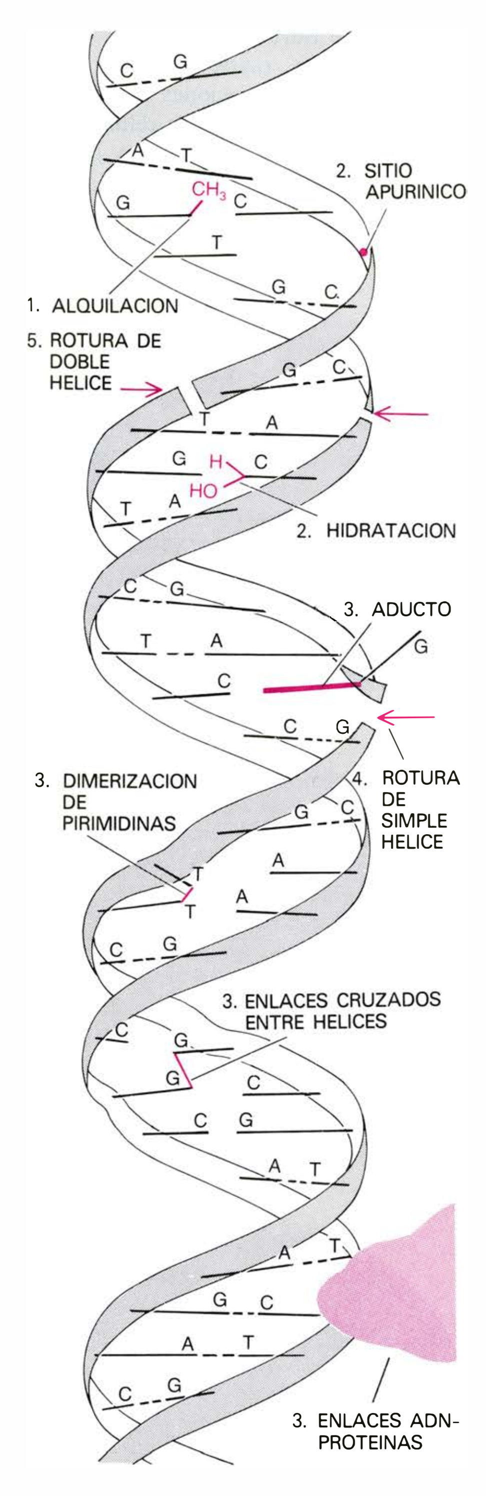
La capacidad cancerígena de los agentes físicos, y especialmente la de las radiaciones ionizantes, se conoce mucho mejor que la de los agentes químicos. La conciencia a nivel popular de que las radiaciones pueden provocar cáncer ha aparecido sólo en la era de las armas y reactores nucleares, pero el peligro ya se

puso de manifiesto poco después del descubrimiento de los rayos X por Wilhelm Konrad Röntgen en 1895. Al cabo de cuatro años se tuvo ya conocimiento de que un técnico que revisaba tubos nuevos de rayos X irradiando su propia mano sufrió cáncer de piel, muriendo a causa de ello. La advertencia fue ignorada, y la mayor parte de la primera generación de radioterapeutas murió de cáncer. El efecto cancerígeno de la radiación ultravioleta de la luz solar se conoce también desde hace algún tiempo. Ya en 1905 un físico francés llamado Dubreuilh observó que el cáncer de piel de la parte posterior del cuello era especialmente frecuente entre los trabajadores que quedaban expuestos al sol al laborar los viñedos y vendimiar la uva de la región de Burdeos.

Aunque los rayos X, los rayos gamma y otras radiaciones inciden sobre todos los componentes celulares de forma aleatoria, se admitió muy pronto que el material genético debía ser el blanco celular más sensible a la radiación. El ADN constituye sólo una pe-



TRES ENSAYOS BACTERIANOS de carcinógenos potenciales revelan la lesión del ADN: el test de Ames (*arriba*), el inductest (*en medio*) y el mutatest lambda (*abajo*). En cada caso, la placa de cultivo de la izquierda es un control sin tratar; las bacterias de ensayo de la placa central fueron tratadas con una dosis moderada y las de la derecha con una dosis superior. El test de Ames muestra la cantidad de mutaciones reversas de *Salmonella typhimurium* deficiente en histidina, que permite la proliferación de las bacterias que han sufrido esta reversión. A la izquierda se observa un fondo de colonias originadas por mutaciones espontáneas (*color rojo*). Crecen muchas más bacterias cuando las bacterias de ensayo son expuestas a 250 nanogramos (*centro*) y 750 nanogramos (*derecha*) del potente mutágeno (y carcinógeno) nitrosoguanidina. En el inductest (*en medio*) la lesión del ADN se pone de manifiesto por la inducción de un profago, un virus bacteriano latente integrado en el ADN de una *Escherichia coli* "lisogénica": las partículas de fago maduras lisan las bacterias de ensayo y forman halos de lisis en una capa de bacterias indicadoras de la cepa A (*color rojo*). En este caso, el deterioro del ADN fue causado por el fármaco antitumoral mitomicina C, 10 nanogramos en la placa central y 200 en la de la derecha. En el mutatest (*abajo*), una forma modificada del profago forma placas en una capa de *E.coli* de la cepa *B* (en la que los fagos no mutados no pueden formar placas) cuando sufren mutaciones en sus regiones "operadoras" y no pueden persistir, a consecuencia de este hecho, en su estado de latencia. También aquí el tratamiento fue de 10 (*centro*) y 200 (*derecha*) nanogramos de mitomicina C. (Los ensayos convencionales resultan lentos y muy costosos.)



LAS RADIACIONES o ciertos agentes químicos provocan alteraciones estructurales en la doble hélice de ADN, dos cadenas de azúcar y grupos fosfato (cintas helicoidales) unidas por bases apareadas: adenina (*A*) y timina (*T*) o guanina (*G*) y citosina (*C*). Las alteraciones pueden encuadrarse en cinco categorías, de las que se ilustran ejemplos: distorsiones insignificantes de la hélice, como la alquilación de una de las bases; distorsiones menores (2) provocadas por la hidratación o ausencia de una base; distorsiones mayores (3) causadas por la inserción de un aducto, la unión de dos bases formando un dímero, o la unión entrecruzada de dos filamentos o de un filamento y una proteína; roturas de un solo filamento (4) o de los dos (5). Cualquier alteración estructural afecta a la función del ADN como molde durante la duplicación pero algunas alteraciones despreciables no son detectadas como lesión del ADN.

queña fracción de los componentes químicos de la célula, pero las lesiones que sufre, directa o indirectamente, son de gran trascendencia para el futuro de una célula, mientras que el deterioro de las proteínas u otros componentes celulares tiene mucho menos efecto. Ello obedece a que el ADN es la memoria de la célula. El ADN se duplica engendrando ADN, de forma que los errores del ADN se transmiten de una generación celular a otra.

La doble hélice de ADN consta de dos cadenas de azúcar (desoxirribosa) y grupos fosfato, enlazadas entre sí por bases nitrogenadas que recuerdan los peldaños de una escalera de caracol: dos purinas (adenina y guanina) y dos pirimidinas (timina y citosina). La adenina se aparea siempre con la timina y la guanina con la citosina, pero la secuencia de bases a lo largo de un filamento de ADN es variable y contiene un determinado mensaje codificado. Cualquier alteración de la estructura de la doble hélice, por pequeña que sea, afecta las funciones del ADN, una de las cuales es servir de molde para su propia duplicación.

No todas las alteraciones del ADN son apreciadas por la célula como una lesión del ADN. Definida con precisión, una lesión del ADN es una alteración que supone un obstáculo para la maquinaria de duplicación y, por tanto, impide la duplicación del ADN poniendo en peligro la supervivencia de la célula. Una vez que se ha producido, la lesión del ADN ha de ser reparada, lo que se lleva a cabo mediante la colaboración de un elevado número de enzimas, codificados por genes, que, en consecuencia, son responsables de la reparación. La reparación nunca es totalmente eficaz, por lo que muchas células mueren. No obstante, puede darse el caso de que algunas células sobrevivan incluso sin haber sido totalmente eliminadas las lesiones del ADN, debido a que un proceso particular de reparación habrá hecho posible la duplicación del ADN lesionado. Entonces, aunque quedará reconstruida la estructura de la doble hélice, el mensaje codificado será distinto del original. Las cicatrices que quedan en el ADN tras la actuación de este proceso son mutaciones.

La correlación entre el cáncer inducido por radiación y las lesiones del ADN inducidas por radiación se había hecho patente a los radiobiólogos desde hace tiempo, pero sólo recientemente se ha hallado una demostración a nivel molecular de que la lesión del ADN es una causa directa del cáncer. La prueba proviene de los pacientes que sufren xeroderma pigmentoso, que son extraordinariamente sensibles a la luz del sol y aun siendo muy jóvenes padecen cáncer de piel, a causa del cual pueden llegar a morir. James E. Cleaver, de la Facultad de Medicina de la Universidad de California en San Francisco, y Dirk Bootsma, de la Universidad Erasmo de Rotterdam, han demostrado que los sujetos pacientes de xeroderma sufren un defecto muy bien descrito: sus células no pueden llevar a cabo un determinado proceso de reparación del ADN.

En un mediodía soleado el flujo de radiación ultravioleta que alcanza la Tierra es suficientemente elevado como para provocar la formación de dímeros de pirimidina en el ADN de las células expuestas, por unión lateral de dos timinas adyacentes o de una timina y una citosina. La mayoría de estas lesiones del ADN de las células epidérmicas se reparan, en las personas normales, a través de un proceso de escisión sorprendentemente similar al de "escisión-reparación" ("cut and patch") que tiene lugar en las bacterias expuestas a la misma radiación. En los pacientes de xeroderma, las lesiones permanecen sin reparar y su acumulación provoca la transformación maligna de las células: las lesiones del ADN engendran el cáncer.

El efecto cancerígeno de determinados productos químicos presentes en el medio ambiente se han ido apreciando lentamente, aun cuando se conozcan desde hace tiempo ejemplos de cáncer causados por lo que podríamos llamar una exposición profesional. Ya en 1775 el médico británico Sir Percival Pott correlacionó la incidencia de cáncer de escroto en hombres que habían trabajado como deshollinadores con la acumulación de hollín en la ingle, ocurrida muchos años antes. Los descubrimientos experimentales sobre carcinógenos químicos se remontan al menos a 1918, cuando dos investigadores japoneses, K. Yamagiwa y K. Ichikawa, mostraron que el cáncer de piel podía provocarse mediante aplicaciones repetidas de alquitrán de hulla sobre la piel de la oreja de conejos. Uno de los productos responsables de este tipo de cáncer es el benzo [*a*] pireno, presente también en el hollín, el humo de los cigarrillos y la carne chamuscada.

El progreso hacia la comprensión de la carcinogénesis química ha sido lento debido al menos a tres razones. Primero, la mayoría de los carcinógenos químicos no son biológicamente activos en su forma original, de manera que, al analizarlos en dicho estado, no se descubre su naturaleza cancerígena. Desde hace sólo una década ha quedado claro, en

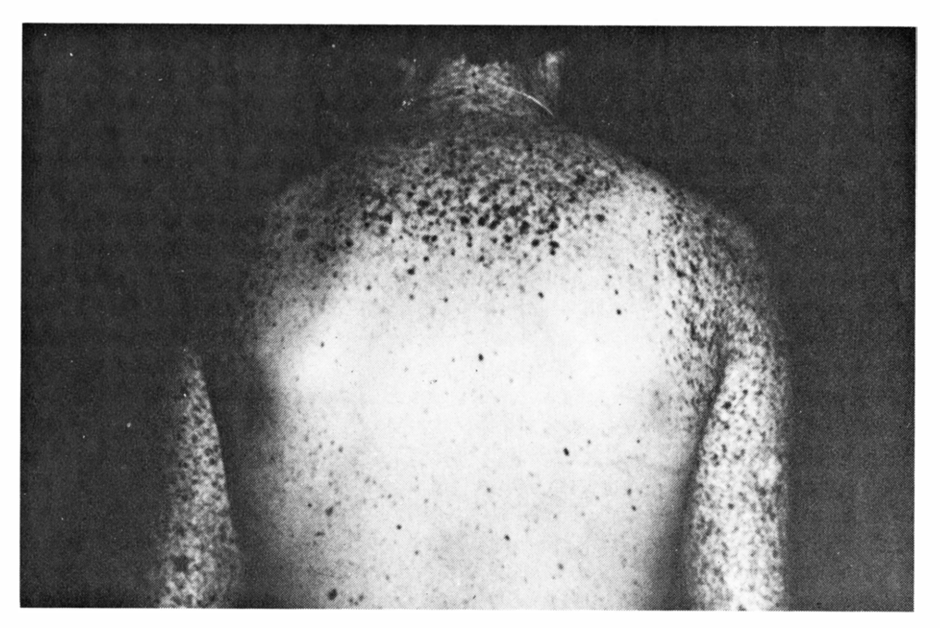
particular como resultado de las investigaciones de James A. Miller y Elizabeth C. Miller, de la Universidad de Winsconsin, que los procesos metabólicos normales, que transforman los alimentos en substancias que el organismo puede absorber y eliminar, y convierten los compuestos dañinos en inofensivos, transforman ciertas substancias químicas del medio ambiente en metabolitos cancerígenos. Había, pues, que describir estos metabolitos antes de poder considerar como una amenaza los productos de los que provienen.

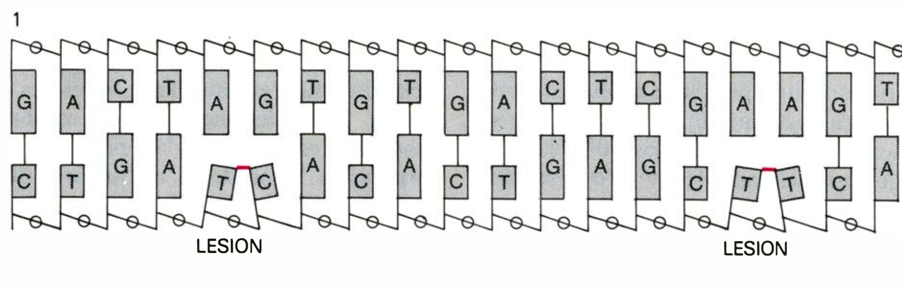
Una segunda causa de retraso fue que los metabolitos carcinógenamente activos reaccionan con diversos componentes celulares, entre ellos el ARN, las proteínas y el ADN. Dado que un gran número de funciones importantes dependen de las proteínas, se creyó a menudo que a la lesión de éstas correspondía el papel más importante en la transformación cancerosa iniciada por cancerígenos químicos. La función clave que desempeña la lesión del ADN en la carcinogénesis por productos químicos se ha descubierto recientemente, y los resultados de los ensayos con bacterias han proporcionado una prueba contundente (aunque indirecta) de dicho mecanismo.

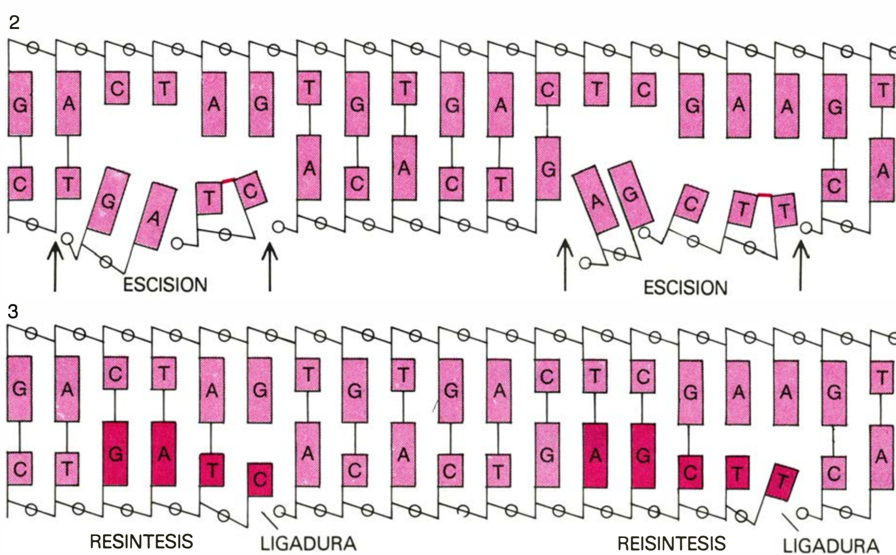
Finalmente, la comprensión de la carcinogénesis química ha estado enmascarada por el hecho de que la lesión del ADN, aun siendo esencial en la iniciación de la carcinogénesis, no provoca normalmente por sí sola la transformación cancerosa: se requieren unos factores adicionales para promover la compleja secuencia de acontecimientos celulares que culminan en la transformación. Los agentes que dañan el ADN son, por tanto, únicamente carcinógenos en potencia.

Como señalé al comienzo de este artículo, sólo una pequeña fracción del torrente de productos químicos que llegan al mercado anualmente puede analizarse con precisión por medio de los habituales ensayos con animales. A fin de obtener resultados con significación estadística debe (o cuando menos debería) someterse a prueba un gran número de animales por cada producto analizado; por este exclusivo motivo sería ya impracticable el análisis a gran escala de productos nuevos (salvo unos pocos, como los aditivos alimenticios, los medicamentos y los productos cosméticos, en los que es obligatorio hacerlo). Aun siendo factibles las pruebas con animales, los fabricantes necesitan tests rápidos y económicos si deben identificar los compuestos que deterioran el ADN cuando los nuevos productos están todavía en elaboración, y se está a tiempo aún de buscar substancias alternativas.

El estudio de la epidemiología del cáncer humano ha proporcionado gran información sobre carcinógenos ambientales, pero estos estudios tienen poco valor inmediato en la detección de nuevos carcinógenos potenciales, ya que la mayoría de cánceres humanos aparecen al cabo de unos 20 o 30 años después de la exposición al agente que los ha provocado. El análisis epidemiológico ha tenido éxito sobre todo en la detección de productos químicos a los que están expuestas subpoblaciones identificables, como las de trabajadores de un







XERODERMA PIGMENTOSO es una enfermedad hereditaria cuyas víctimas sufren cáncer de piel, habitualmente en las partes no cubiertas del cuerpo, como queda mostrado en la fotografía superior, a consecuencia de una exposición normal al sol. En estos pacientes no se da un proceso de reparación del ADN que en los individuos normales elimina los dímeros de timina, las lesiones creadas por la radiación ultravioleta presente en la luz solar (*1*). En el proceso normal de reparación se desprenden primero las bases del trozo lesionado (*2*). Los segmentos de ADN desprendidos son resintetizados y los nuevos segmentos se unen a segmentos no dañados (*3*). El xeroderma pigmentoso puede causar la muerte.

determinado ramo industrial, y que por ello muestran una elevada incidencia de algún tipo de cáncer. Para ciertos carcinógenos extendidos ampliamente no hay subpoblaciones identificables que se presten al análisis estadístico.

No cabe duda alguna de que la presente situación requiere métodos simples, rápidos y económicos de detección de carcinógenos potenciales, es decir, de agentes que lesionan el ADN. Una posibilidad es medir el daño del ADN directamente en células de mamíferos, determinando bioquímicamente la incidencia de formas específicas de lesión del ADN en células expuestas a un producto químico. Tales estudios los están llevando a cabo biólogos moleculares y sus resultados proporcionan valiosos criterios de referencia. A efectos de análisis en masa, no obstante, resulta más rápido y económico detectar y medir la magnitud de la lesión del ADN registrando sus manifestaciones en bacterias.

Una de las grandes ventajas de los análisis efectuados con bacterias es la enorme amplificación biológica implícita en toda manipulación de bacterias. Es sencillo hacer crecer unos mil millones (10^9) de bacterias por mililitro de medio de cultivo. Una mutación como el cambio de un solo par de bases en el ADN bacteriano, que es imposible detectar mediante los métodos bioquímicos al uso, quedaría puesta de relieve en forma de una bacteria mutante. Esta puede ser seleccionada de entre las 10^9 células ya que sus células hijas, y sólo ellas, pueden proliferar y dar lugar a una colonia apreciable a simple vista en una placa de agar nutritivo. Dado que una colonia consta de alrededor de un millón (10^6) de bacterias, una mutación poco frecuente que tenga una probabilidad de, por ejemplo, uno entre 100 millones (una probabilidad de 10^{-8}) quedaría amplificada por un factor de 10^{14} (cien billones).

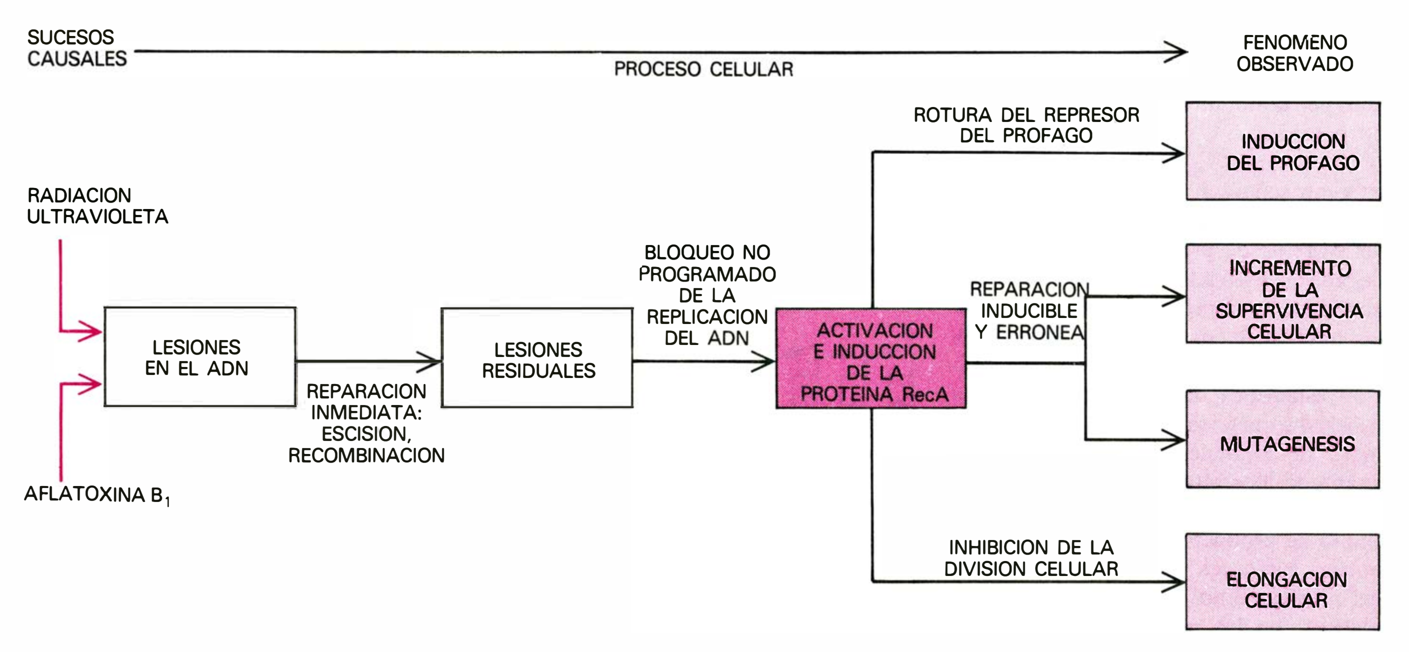
La serie de acontecimientos celulares que conducen a una mutación, o a otras manifestaciones de deterioro del ADN en las bacterias ha sido, hasta cierto punto, clarificada recientemente. La bacteria *Escherichia coli*, que habita en el colon de numerosos mamíferos, seres humanos incluidos, está programada genéticamente para dividirse y, bajo condiciones óptimas de crecimiento, formar dos bacterias hijas cada 30 minutos. Cuando el ADN de una bacteria ha sufrido algún daño, la bacteria necesita más de dos horas para completar su ciclo de división (si no ha muerto a causa de la lesión). Durante este tiempo tiene lugar una sucesión de fenómenos celulares.

Inmediatamente después de producirse las primeras lesiones del ADN, se lleva a cabo una primera operación de reparación: un proceso de escisión en el que los segmentos dañados son separados del ADN y reemplazados por una secuencia de ADN no deteriorada. Pueden, sin embargo, quedar lesiones residuales. Si las lesiones remanentes son grandes o están agrupadas, pueden entorpecer la labor de la maquinaria de duplicación del ADN; la duplicación se detiene entonces bruscamente. A menos que la interrupción de la duplicación sea transitoria, la supervivencia de la célula está en peligro. Como sugirieron por primera vez Miroslav Radman, de la Universidad Libre de Bruselas, y Evelyn M. Witkin, de la Universidad de Rutgers, se ponen en marcha entonces los siguientes mecanismos celulares de adaptación destinados a superar esta situación difícil o de emergencia:

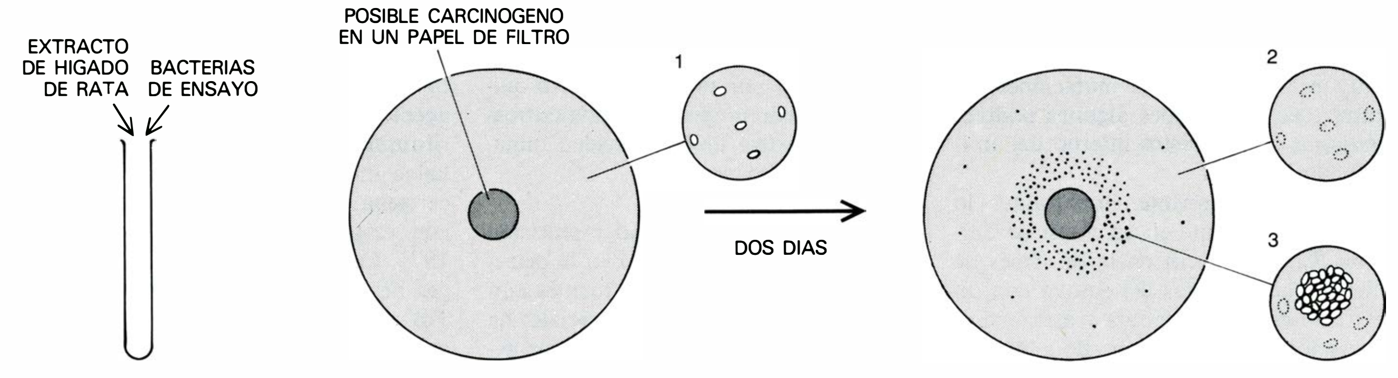
1. Se pone en marcha una operación de reparación. Dicho proceso de reparación inducible y propenso a cometer errores (al que se denomina "reparación SOS") tiende a restaurar la estructura del ADN aunque se produzcan errores en el mensaje codificado; de hecho esta respuesta adaptativa puede que haya tenido éxito, en parte porque no impide que se sigan las leyes de la duplicación normal del ADN. En todo caso, el precio pagado por la supervivencia celular parece ser con bastante frecuencia la aparición de mutaciones.

2. Cesa la división celular. La elongación de las células, que ordinariamente precede a la división, se acentúa y las células pueden llegar a formar filamentos. Puede que la elongación sea adaptativa, ya que facilita la recombinación entre las dos dotaciones de cromosomas dañados de la célula, pudiendo combinarse los segmentos intactos para dar lugar, al menos, a un cromosoma funcional.

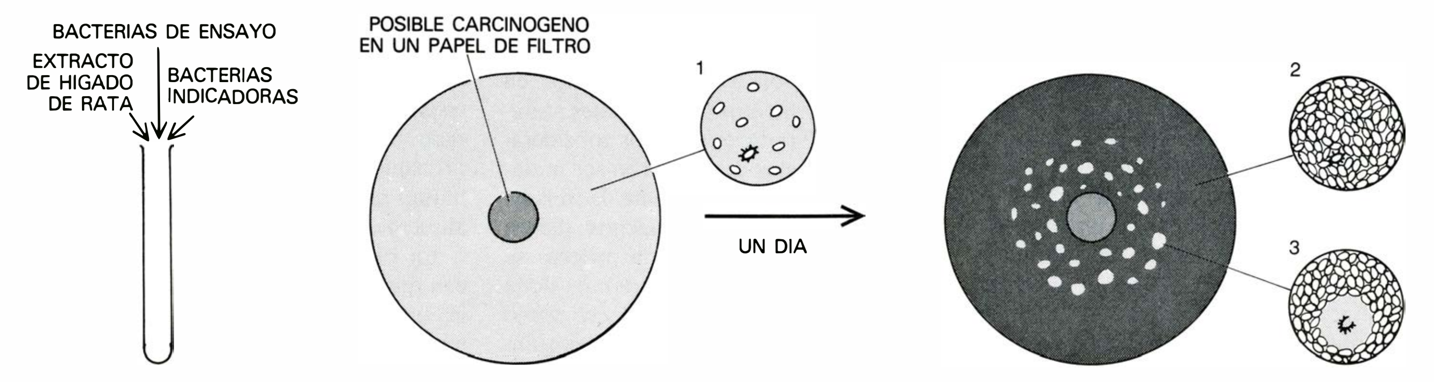
3. Si la bacteria contiene un profago, es decir, el ADN de un virus bacteriano en estado aletargado, dicho profago es inducido a multiplicarse, dando lugar a un elevado número de fagos maduros que salen de la célula lisándola. Esta res-



SECUENCIA DE FENOMENOS CELULARES que siguen a la lesión del ADN provocada en *E.coli* por la radiación ultravioleta o por el carcinógeno aflatoxina B1. Los procesos de reparación inmediata corrigen la mayoría de las lesiones pero dejan un deterioro residual que provoca una detención imprevista de la duplicación. Esta amenaza contra la supervivencia de la célula activa la proteína RecA transformándola en un enzima degradador de proteína, e induce asimismo la síntesis de más RecA. Las diferentes formas de la proteína están involucradas en tres procesos (*derecha*).



MUTAGENESIS que se detecta en el test de Ames mezclando un extracto de hígado de rata (que suple las funciones metabólicas de los mamíferos) con bacterias de ensayo (que no pueden proliferar debido a que una mutación les impide sintetizar histidina, un nutriente indispensable) y sembrando la mezcla en un medio de cultivo de agar, de forma que una fina capa de bacterias cubra el medio, como se muestra a escala microscópica (*1*). En esta prueba se coloca una dosis del compuesto a ensayar en un disco de papel de filtro sobre las bacterias. Al cabo de dos días, la mayoría de las bacterias *his*$^-$ han muerto por falta de histidina (*2*), pero el daño causado al ADN por el compuesto químico que ha difundido del disco ha provocado mutaciones, algunas de las cuales dan lugar a una reversión de la mutación *his*$^-$. Las bacterias que sintetizan histidina se multiplican y colonizan (*3*).



EN EL INDUCTEST se detecta la inducción de un virus bacteriano aletargado, el profago lambda. Las bacterias de ensayo lisogénicas se mezclan con un extracto de hígado de rata y seguidamente con bacterias indicadoras entre las que hay unas cuantas bacterias lisogénicas (*1*). Al cabo de dos días la mayor parte de la placa está cubierta de una gruesa capa de bacterias indicadoras (*2*). En aquellos puntos en que el compuesto sometido a análisis ha difundido del disco de papel de filtro, la lesión del ácido desoxirribonucleico provocada por él da lugar a la inducción de fagos lambda maduros. Las partículas del fago lisan las células bacterianas lisogénicas, salen de ellas y matan las bacterias indicadoras vecinas, dando halos visibles (*3*).

puesta adaptativa desarrollada por el profago le asegura la supervivencia cuando la célula huésped parece estar condenada a morir: las ratas abandonan el barco que se hunde.

Cada una de estas respuestas adaptativas está promovida en parte por una proteína multifuncional denominada RecA (de "recombinación", ya que la lesión de esta proteína perjudica en general la recombinación). Según parece, la proteína RecA viene postulada por *E.coli*, no sólo para la recombinación, sino también para las respuestas enumeradas anteriormente. Cuando queda bloqueada la duplicación del ADN, la proteína RecA se activa y crece espectacularmente su síntesis.

De entre las distintas manifestaciones bacterianas de la lesión del ADN, Ames eligió la mutagénesis como base de su labor pionera en el desarrollo de un ensayo de carcinógenos potenciales. Su prueba *Salmonella*-hígado de mamífero, conocida como el test de Ames, es normalmente el test tomado como criterio, y con mucho, el más utilizado también. El organismo utilizado en la prueba es una cepa de *Salmonella typhimurium*, otra bacteria del colon, portadora de una mutación (*his*$^-$) que le impide fabricar uno de los enzimas requeridos para la síntesis del animoácido histidina, un componente esencial de las proteínas. A consecuencia de la mutación, la bacteria no puede crecer en un medio nutritivo mineral a menos que el medio sea complementado con una fuente externa de histidina.

En muy raras ocasiones la mutación *his*$^-$ sufre una reversión: una mutación reversa restaura la secuencia normal en el código de ADN para el enzima requerido y con ello devuelve el suministro interno de histidina. La reversión se descubre ya que sólo las bacterias que la han sufrido pueden formar colonias en un medio sin histidina. Obviamente, la tasa de mutación reversa, que ordinariamente es muy baja, se incrementará considerablemente si la bacteria *his*$^-$ es expuesta a un producto que provoque mutaciones. Esta es la base teórica del test de Ames.

Ultimamente Ames y sus colegas han debido introducir tres modificaciones importantes en la cepa original *his*$^-$ para hacerla más sensible y adaptable a las condiciones experimentales. Las bacterias como *E.coli* y *S.typhimurium* tienen una cubierta bastante impermeable que reduce o incluso impide la penetración de muchos productos químicos en la célula. (Esta armadura bacteriana ha aparecido ya que las bacterias deben subsistir normalmente en un medio ambiente hostil, como el intestino o las alcantarillas.) Ames y sus colegas superaron la barrera de la pared celular introduciendo una mutación que provocaba defectos en ella. Por otra parte, lograron hacer la cepa más sensible a los agentes que dañan el ADN eliminando su capacidad de llevar a cabo el proceso de reparación por escisión, de manera que la mayor parte de lesiones originales no fuesen reparadas. Por último, introdujeron en la bacteria un plásmido, un elemento genético externo que hace la duplicación del ADN más propensa a errores. Por medio de estas tres modificaciones crearon una cepa en la que tan sólo unas pocas moléculas de carcinó-

geno eran ya capaces de provocar lesiones en el ADN, y estas daban lugar, muy probablemente, a mutaciones; de entre estas mutaciones algunas podrían restaurar el suministro interno del aminoácido histidina.

Lo verdaderamente importante, lo que hizo realmente eficaz el test de *Salmonella*, fue la ocurrencia de Ames de mezclar las bacterias del ensayo con un extracto de hígado de rata y así someter el producto químico analizado a los procesos metabólicos de los mamíferos. Como señalé más arriba, normalmente no es la forma original de un carcinógeno químico la que es activa, sino uno de sus metabolitos. Por ser el hígado la mayor fábrica de metabolitos del organismo, los enzimas de un extracto de hígado de rata deben convertir el producto ensayado en metabolitos que reaccionen con el ADN, en caso de haberlos.

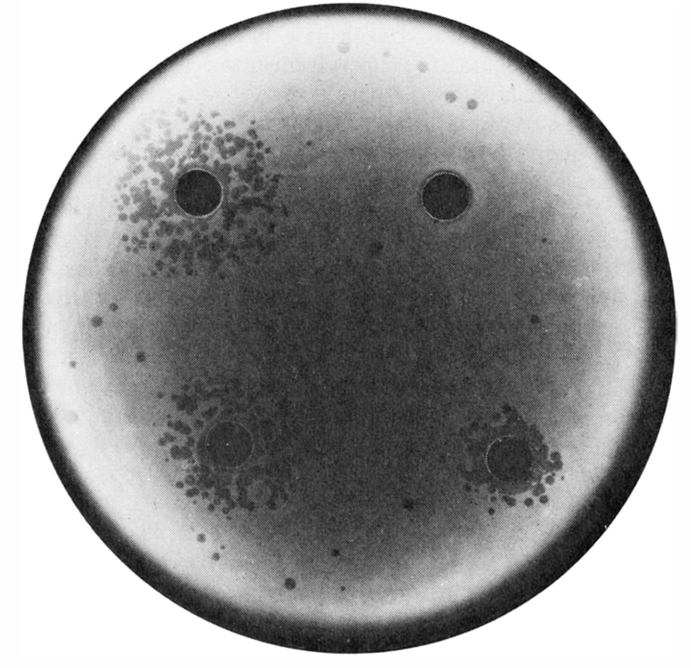
En la práctica, el test de Ames se efectúa normalmente añadiendo el producto a analizar a las bacterias *his*$^-$ del ensayo inmersas en un extracto de hígado de rata y sembrando la mezcla en un medio nutritivo sólido carente de histidina. (Para fines demostrativos, el producto químico puede esparcirse, en cambio, sobre un papel de filtro que se pone entonces en un medio en el que se ha sembrado previamente las bacterias mezcladas con la preparación de hígado.) Tras dos días de incubación, cualquier célula que haya sufrido la mutación reversa dará lugar a colonias. El número de tales colonias por mol de producto químico analizado proporciona una estimación cuantitativa de la capacidad mutagénica de dicho producto.

La simplicidad, sensibilidad y precisión del test con *Salmonella* en la detección de un gran número de fuentes ambientales de carcinógenos potenciales ha dado lugar a que se utilice corrientemente en más de 2000 laboratorios gubernamentales, industriales y universitarios de todo el mundo; se calcula que se han sometido a este ensayo unos 2600 productos químicos. Ames y su colaborador Joyce Mc-Cann han ratificado personalmente la validez del experimento ensayando más de 300 productos que habían sido previamente calificados de cancerígenos o no-cancerígenos en base a experimentos con animales. Alrededor del 90 por ciento de los considerados cancerígenos demostraron ser mutagénicos y aproximadamente la misma proporción de no-cancerígenos dieron un resultado negativo en la prueba. Se han ideado después otros ensayos de la capacidad mutagénica con *E.coli* como bacteria de ensayo, y su eficacia es similar a la prueba original con *Salmonella.*

Pueden mencionarse dos notables logros del ensayo de la capacidad mutágena para dar una idea de su valor. En el Japón se estuvo añadiendo durante algunos años a un gran número de productos alimenticios comunes, como agente antibacteriano, el compuesto furil-furamida, conocido como AF-2. No había mostrado ninguna actividad cancerígena en las pruebas convencionales con ratas en 1962 o con ratones en 1971. En 1973, T. Sugimura y sus colegas del Centro Nacional del Cáncer de Tokyo hallaron que el AF-2 era altamente mutagénico en bacterias: ¡incluso en el aditivo contenido en una sola rodaja de embutido de pescado pudieron comprobar la existencia de actividad mutagénica! El descubrimiento incitó a una nueva serie de ensayos de la capacidad cancerígena con animales, que evidenciaron que el AF-2 era desde luego cancerígeno. Se le retiró del mercado. Si no hubiera sido por el ensayo bacteriano, el AF-2, que había pasado dos tests homologados y había sido declarado no-cancerígeno, seguiría siendo probablemente un componente del embutido de pescado y de otros productos alimenticios japoneses.

En 1975 Ames y sus colegas informaron que el 89 por ciento de los principales tipos de tinte de cabello vendidos en los Estados Unidos contenían compuestos cancerígenos. Desde entonces, la industria de cosméticos ha modificado la composición de la mayoría de los tintes. Se calcula que, sólo en los Estados Uni-





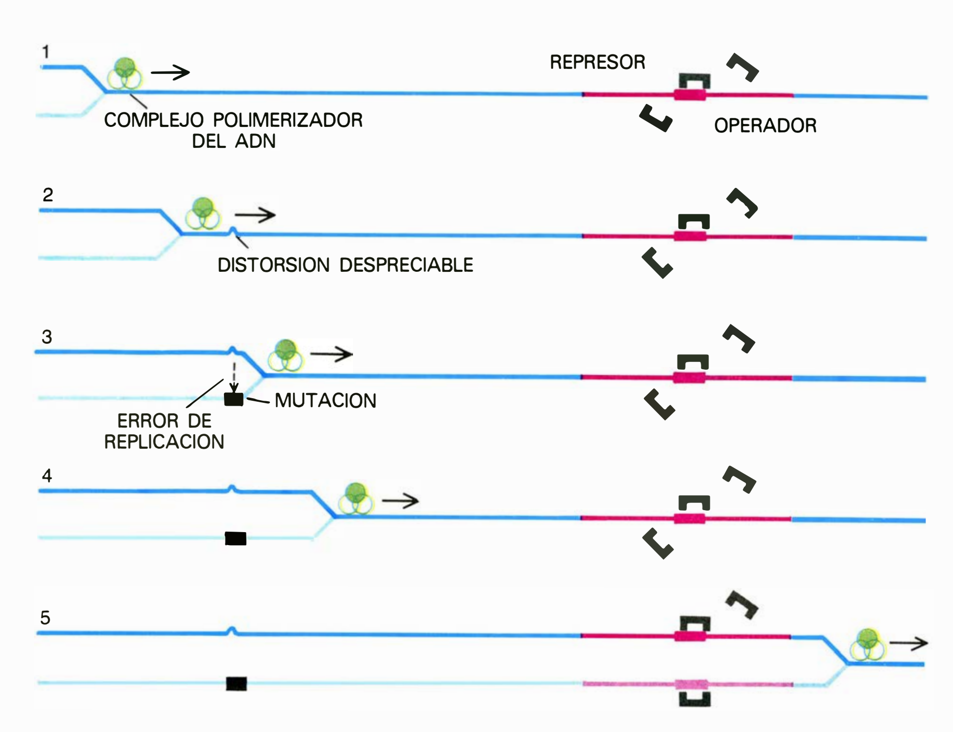
ENSAYOS basados en la formación de "manchas". En ellos se visualiza la eficacia de los ensayos bacterianos cualitativamente, aunque son preferibles los ensayos cuantitativos. En el test de Ames (*izquierda*) el compuesto analizado fue etil-metano-sulfonato, un agente alquilante y potente mutágeno. Alrededor del disco del que ha difundido el mutágeno se aprecia un denso halo de colonias revertientes de *S.typhimurium*. (Cerca del disco hay una zona en que la concentración tóxica del compuesto ha matado todas las bacterias.) Las colonias mayores son bacterias mutantes que han aparecido espontáneamente, sin ser expuestas al compuesto químico. En el inductest (*derecha*) el producto analizado fue aflatoxina B_1, un potente carcinógeno hepático. Las dosis de aflatoxina de los cuatro discos fueron (*en el sentido de las agujas del reloj, desde arriba a la derecha*) 0, 20, 200 y 2000 nanogramos. De fondo se aprecian bacterias de *E. coli* lisogénicas que han producido fagos maduros matando las bacterias indicadoras vecinas, formando halos.

dos, varias decenas de millones de personas se tiñen el pelo, lo que sugiere que los componentes rechazados representaban un riesgo considerable.

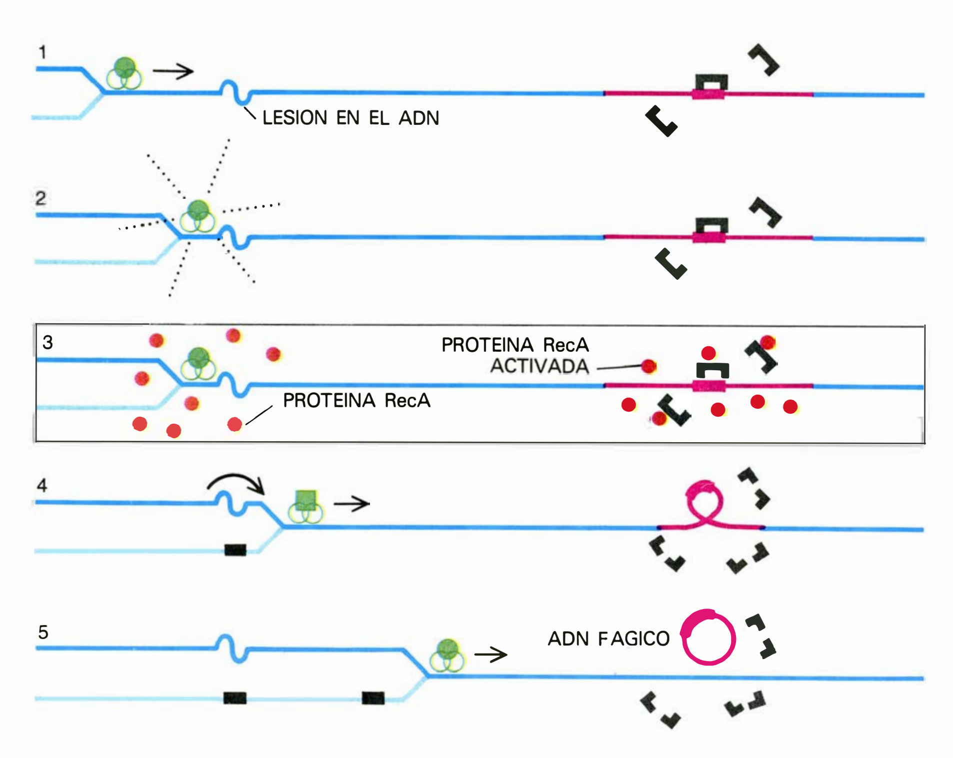
Pese a sus ventajas, los ensayos de la capacidad mutagénica presentan algunas limitaciones técnicas e incluso teóricas. Dado que las mutaciones quedan puestas de relieve en el test de Ames por una restauración de la actividad enzimática, cualquier mutación que no logre reconstruir la secuencia precisa del ADN que codifica para el enzima que sintetiza la histidina, no puede observarse. Por ejemplo, el fármaco antitumoral bleomicina, cuya acción terapéutica consiste en dañar el ADN de las células tumorales (al igual que lo hacen aproximadamente la mitad de todos los agentes antitumorales), no es capaz de inducir la mutación detectada en el test con *Salmonella*. Una respuesta falsamente negativa de este tipo es un problema técnico que puede remediarse sustituyendo las bacterias del ensayo o utilizando otro test de corta duración complementario.

Las respuestas falsamente positivas son más significativas desde un punto de vista teórico, ya que pueden originar dudas sobre la validez de los ensayos de la capacidad mutagénica en la identificación de carcinógenos potenciales. La cuestión es que algunas reacciones químicas con el ADN son altamente mutágenas en bacterias y células de mamíferos sin que sean cancerígenas, que se sepa. Entre ellas está la incorporación de un análogo de una de las bases nitrogenadas y la metilación de ciertos puntos de las bases. Estas reacciones causan alteraciones insignificantes en la estructura del ADN que no son percibidas por la célula como lesión del ADN; la duplicación del ADN se efectúa con normalidad y el nuevo ADN porta un mensaje que puede ser leído incluso si es erróneo. Este proceso se denomina mutagénesis directa, y en la detección de mutaciones debiera distinguirse claramente de la mutagénesis indirecta, más frecuente. En este último proceso, descrito anteriormente, la lesión del ADN comporta una detención transitoria de la duplicación; la duplicación se completa con ayuda de la proteína RecA sobre un molde que contiene lesiones, provocando la aparición de mutaciones en el filamento recién formado.

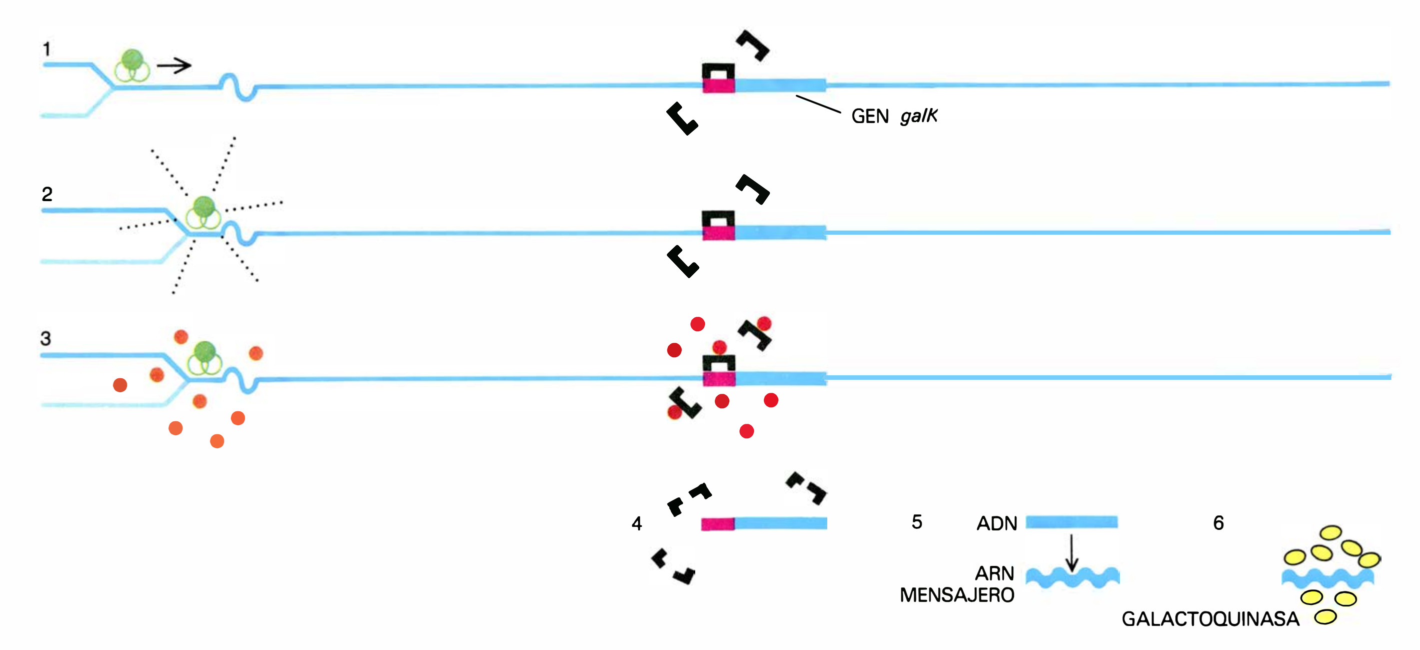
Dado que es el deterioro del ADN lo que aparentemente inicia la transformación cancerosa de las células de los mamíferos, un compuesto es potencialmente cancerígeno si es un agente causal de lesiones en el ADN, no sólo porque provoque mutaciones. Un agente mutágeno puede tener graves efectos ge-



MUTACION DIRECTA, que es el resultado de la duplicación del ADN que contiene una pequeña distorsión en un filamento. Esta figura y la siguiente muestran la duplicación de un solo filamento de ADN bacteriano (*azul*) que contiene un segmento de ADN vírico, el profago lambda (*rojo*), mantenido en un estado de latencia por una proteína represora que bloquea una región operadora. El filamento de ADN es duplicado por un complejo polimerizante (*1*). Una distorsión inapreciable del ADN (*2*) no obstruye la duplicación, pero da lugar a un error en la duplicación, de manera que el filamento de ADN recién sintetizado (*azul claro*) contiene una mutación (*3*). La diminuta distorsión no se aprecia como lesión del ADN, la proteína RecA no es inducida, el represor se mantiene y el profago se duplica en su estado de latencia (4,5). El deterioro del ADN inicia la transformación cancerosa de las células.



MUTACION INDIRECTA, que es el resultado de los fenómenos que suceden a la lesión del ADN (*1*) que bloquea la duplicación y produce una señal de "SOS" (*2*). Se sintetiza una gran cantidad de proteína RecA (*naranja*) (*3*). La proteína RecA facilita de alguna forma la duplicación del ADN sobre un molde deteriorado, en apariencia modificando temporalmente el complejo polimerizante (*4*). Esta duplicación es altamente mutágena, no sólo en el lugar de la lesión sino también en otros puntos (*5*). La proteína RecA activada degrada el represor (4, *derecha*), liberando la región operadora del ADN del profago. De esta forma, el profago es inducido a formar ácido desoxirribonucleico vírico (*5*), que a su vez, produce partículas víricas que abandonan la bacteria huésped lisándola, esto es, destruyéndola.



TEST DE INDUCCION ENZIMATICA, que es una contrapartida bioquímica al inductest. El gen *galK*, que dirige la síntesis de la galactoquinasa, se inserta en el ADN del profago lambda de tal forma que la síntesis del enzima queda bloqueada por el represor de lambda. Al resultar dañado el ADN, la proteína RecA activada degrada el represor. El ADN *galK* liberado (*4*) se transcribe a ARNm (*5*), el cual se traduce a galactoquinasa (*6*).

néticos sobre una población biológica al alterar la información codificada en el ADN de las células germinales de los individuos, y aun así no ser capaz de provocar el cáncer ya que no daña realmente el ADN de las células somáticas: las células del resto del cuerpo. No todas las substancias genéticamente tóxicas (compuestos que afectan al ADN) son potencialmente cancerígenas, mientras que todos los carcinógenos químicos son mutágenos indirectos ya que lesionan el ADN. En otras palabras, la actividad mutagénica en las bacterias está correlacionada con la carcinogénesis en los mamíferos principalmente por la mutagénesis indirecta que proviene de las lesiones del ADN.

En 1953, antes siquiera de que se hubiera descrito la estructura del ADN, André Lwoff, del Instituto Pasteur de París, anticipó que "las bacterias lisogénicas inducibles podrían ser un buen medio de detección de la actividad cancerígena, e incluso de la anticancerígena". Se dice que una bacteria es lisogénica si contiene, en un estado de latencia, el ADN de un virus bacteriano "temperado", que en este estado latente se denomina profago. Uno de tales virus temperados es el fago lambda, que se transforma en profago cuando se integra en el ADN de ciertas cepas lisogénicas de *E.coli*.

Cuando las bacterias lisogénicas *E.coli* se someten a cualquier tratamiento que detiene la duplicación del ADN, se activa o induce el profago: su ADN se separa del ADN bacteriano, forma un anillo, dirige la síntesis de las proteínas que componen la partícula vírica y se desarrolla la descendencia de fagos maduros provocando la lisis de la célula huésped y saliendo de ella. El proceso se denomina inducción lisogénica. Bajo condiciones normales, el estado latente del profago se mantiene por un represor, una proteína acoplada a las regiones "operadoras" del ADN y que bloquea el funcionamiento de los genes de lambda (salvo el gen que dirige la síntesis del represor que mantiene latente al profago). Jeffrey W. Roberts y Christine W. Roberts, de la Universidad de Cornell, descubrieron que la inducción del profago lambda resulta de la degradación del citado represor; junto con Nancy Craig han demostrado recientemente que el represor relativamente purificado puede ser degradado al mezclarlo con una forma activada de la proteína RecA.

Provocando la activación de la proteína RecA, es decir, llevándola a la forma en que puede degradar el represor vírico, la lesión del ADN da lugar a la inducción del virus latente. La inducción del profago lambda puede, pues, servir de test de la lesión del ADN. Aun antes de haber comprendido a nivel molecular el mecanismo de desarrollo del profago, se empleó la inducción lisogénica en la industria farmacéutica para identificar posibles antibióticos y medicamentos antitumorales. No obstante, al aplicar los ensayos de inducción con cepas ordinarias de *E.coli* lisogénica no se obtuvo una respuesta positiva para carcinógenos ya conocidos, como el benzo[*a*]pireno.

Patrice Moreau, Adriana Bailone y yo nos propusimos hace cinco años renovar el test de inducción de lambda para hacerlo capaz de la identificación de cualquier agente que lesionara el ADN. Nos hicimos el razonamiento siguiente: Puesto que el mecanismo molecular de la inducción lisogénica se conocía mejor que el de la mutagénesis, un test de inducción del profago proporcionaría más detalles sobre el efecto concreto de los carcinógenos químicos sobre el ADN que los ensayos de mutagénesis, y sería también una herramienta suplementaria para la detección en masa.

Siguiendo las directrices trazadas por el trabajo de Ames, utilizamos bacterias de ensayo permeables a los productos químicos y deficientes en enzimas del proceso de reparación por escisión, y ensayamos carcinógenos potenciales en presencia de una mezcla metabolizante de hígado de rata. El ensayo de inducción del profago, o inductest, ha logrado discriminar de manera efectiva entre carcinógenos y no-carcinógenos en nuestro laboratorio y en muchos otros.

El inductest tiene ciertas ventajas sobre los tests de mutagénesis. En éstos, las mutaciones pueden detectarse sólo en bacterias que han sobrevivido a la exposición de un tratamiento químico que a menudo es tóxico, además de mutágeno, y no es probable que más de una bacteria de cada 1000 pueda ser detectada, por haber sufrido una reversión de la mutación *his*⁻. Por el contrario, la inducción lisogénica puede observarse en cualquier bacteria de la población, pues la inducción del profago es un fenómeno que alcanza a todas las bacterias y no depende prácticamente de su supervi-

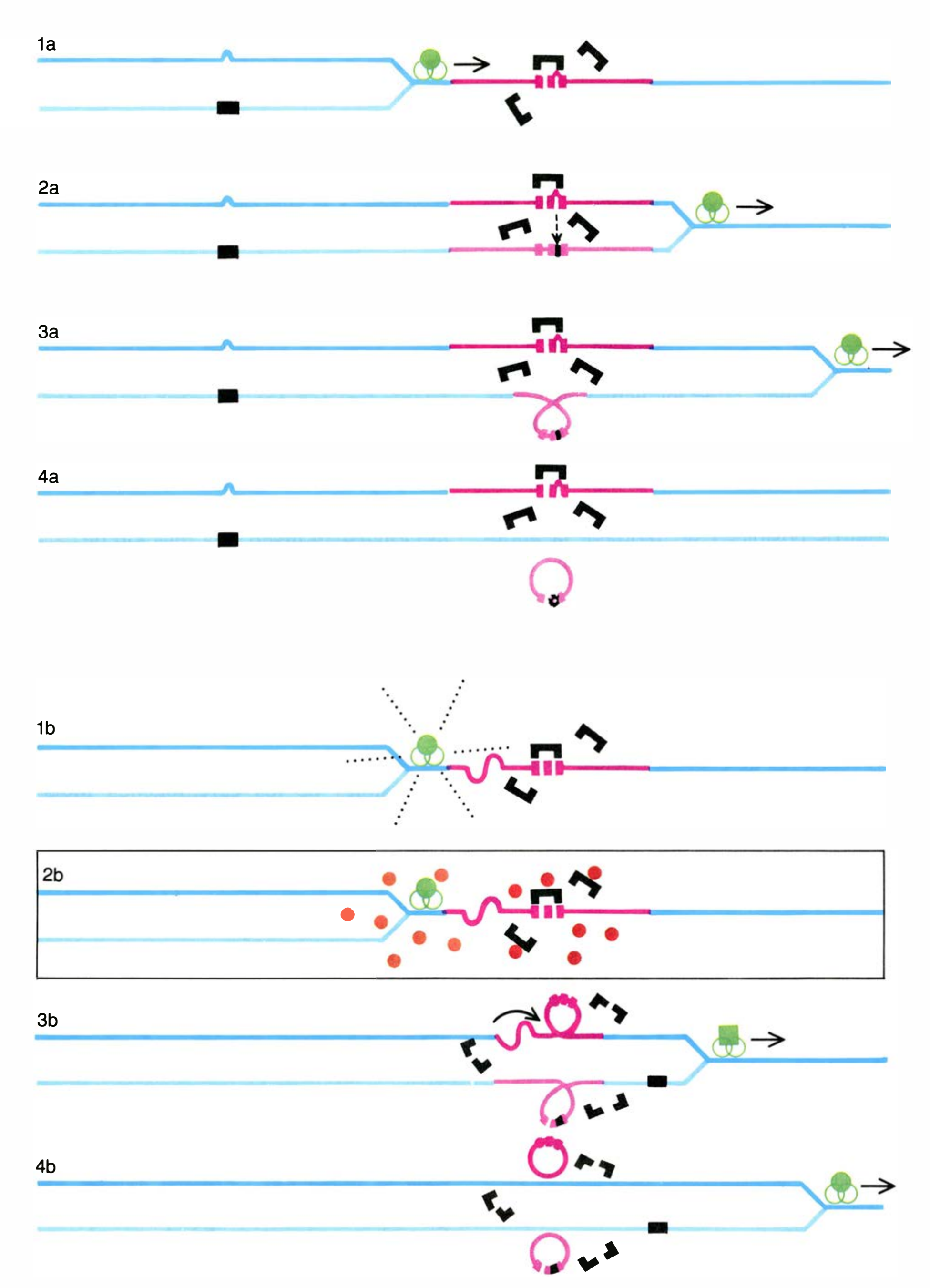
vencia al tratamiento químico; en cualquier caso, una bacteria lisogénica que es inducida y produce una descendencia de fagos muere siempre. El inductest, por tanto, sigue dando respuesta positiva para un carcinógeno potencial altamente tóxico a dosis que matarían las bacterias del ensayo en un test de mutagénesis.

El hecho de que en el inductest la mayoría de las células sufran un cambio espectacular, nos condujo a diseñar un ensayo bioquímico de la inducción lisogénica. Sankar Adhya, Maxwell E. Gottesman y Asis Das, del Instituto Nacional del Cáncer de los Estados Unidos, construyeron una cepa bacteriana en la que el gen que codifica para el enzima galactoquinasa había sido integrado en el ADN de lambda de forma que el represor de lambda bloquease la síntesis del enzima. Alain Levine, Moreau, Steven Sedgwick y yo demostramos que cuando un compuesto químico activa la proteína RecA de esta cepa, el represor es degradado y se sintetiza galactoquinasa, y que la cantidad de actividad enzimática detectada permite evaluar el número de lesiones producidas en el ADN. Este ensayo bioquímico puede que sea la más breve de las pruebas de corta duración: sólo se tarda medio día en analizar un carcinógeno potencial.

Moreau y yo hemos construido también cepas que contienen una nueva forma de profago lambda con las que pueden identificarse compuestos que inducen la aparición de mutaciones sin dañar el ADN. Este "mutatest lambda", como lo denominamos nosotros, pone de manifiesto si un producto químico da lugar a mutaciones directamente (es decir, sin dañar el ADN y, por lo tanto, sin inducir al profago) o indirectamente (deteriorando el ADN y produciendo virus activos).

Aun cuando ni el inductest ni el mutatest lambda han sido ratificados todavía tan ampliamente como el ensayo de la capacidad mutágena de Ames, se puede ya apreciar su interés y el complemento que aportan a los ensayos de mutagénesis: ellos permiten, en particular, determinar la naturaleza de las lesiones sufridas por el ADN y establecer una correlación entre la capacidad de dañar al ADN de bacterias y la de provocar el cáncer en mamíferos. Quizás, a la larga, resulte más importante un sistema biológico que comprenda tanto células como virus latentes ya que podría ser una valiosa herramienta de análisis de las vías de las transformaciones bioquímicas, genéticas o no (degradación del represor, por ejemplo), provocadas por la exposición de las células a carcinógenos.

Dado que el hombre está expuesto a un torrente de productos químicos, de los que algunos aumentan el riesgo de cáncer, hay una necesidad urgente de identificar carcinógenos potenciales y disminuir la exposición de los seres humanos a ellos hasta un nivel de riesgo despreciable, y mantener dicho nivel. En este empeño, los ensayos bacterianos pueden representar un papel fundamental. Con vistas a asegurar una aceptación general de criterios uniformes, algunos investigadores han favorecido intensamente la adopción de una prueba de mutagénesis única, estandarizada. Sin embargo, a menudo es aconsejable hacer un meticuloso estudio de las propiedades físicas y químicas particulares de la substancia en cuestión. Parece, pues, más sensato someter cada compuesto



MUTAGENESIS DIRECTA E INDIRECTA. Pueden distinguirse entre sí por medio del mutatest, basado en el mecanismo aquí representado. El represor se une a un operador levemente modificado (*1a, 1b*). Una alteración diminuta del ADN operador (*1a*) da lugar, durante la duplicación, a una mutación (*2a*) que impide ya la unión del represor (*3a*). El profago forma partículas de fago maduras (*4a*) que quedan puestas de manifiesto selectivamente por la cepa bacteriana indicadora *B*, no sensible a fagos no mutados; este desarrollo de fagos, que ocurre independientemente de la activación de la proteína RecA, revela la mutagénesis directa. Por otro lado, la lesión del ADN provoca, durante la duplicación (*2b, 3b*) una mutagénesis indirecta dependiente de la proteína RecA, detectada por la cepa de bacterias indicadoras *B*. El acontecimiento concomitante de la inducción del profago (*3b, 4b*) resultante de la activación de la proteína RecA se detecta con las bacterias indicadoras de la cepa *A*, no con las de la cepa *B*. La **mutados; este desarrollo de fagos, que ocurre independientemente de la activación de la proteína RecA, revela la mutagénesis directa. Por otro lado, la lesión del ADN provoca, durante la duplicación (*2b, 3b*),**

químico a una batería de tests. Todo test tiene sus puntos débiles, que dan lugar a respuestas falsamente positivas o negativas. Diversas determinaciones independientes y complementarias de la capacidad de lesión del ADN de un producto pueden proporcionar una evaluación compensada y precisa del riesgo.

El reconocimiento de que un determinado producto químico es potencialmente cancerígeno tiene implicaciones sociales, económicas e incluso políticas obvias. El compuesto puede representar un riesgo a una gran población, como fue el caso del aditivo alimenticio AF-2 en el Japón. La situación pudo resolverse fácilmente: el AF-2 pudo ser sustituido por un producto inocuo. No era imprescindible y bastó con retirarlo del mercado. Otros compuestos cancerígenos no pueden eludirse con tanta facilidad y no pueden eliminarse del mercado; han surgido problemas con substancias tan distintas como el humo de los cigarrillos, el de los vehículos a motor y los fármacos antitumorales. En el caso de los carcinógenos más imprescindibles es claramente necesario reducir la exposición humana hasta niveles lo más seguros posible. Esto requiere la adopción de un conjunto ampliamente aceptado de normas de seguridad apoyadas en una legislación rigurosa.

Estamos hoy en día, en lo que se refiere a los carcinógenos químicos, donde estábamos hace tres décadas respecto de las radiaciones ionizantes. La llegada de los reactores de energía nuclear hizo necesario proteger a los trabajadores de la industria de fuentes directas de radiación y defender a cualquier persona de la exposición a líquidos radiactivos o desechos sólidos. Un órgano internacional auspiciado por las Naciones Unidas desarrolló un conjunto de normas y reglamentos diseñados para reducir al mínimo la exposición humana a distintas fuentes de radiación. Las normas, ampliamente aceptadas, se pusieron en práctica y han resistido el paso del tiempo. Estas reglamentaciones deben proporcionar valiosas fuentes de referencia en el desarrollo de criterios para los carcinógenos químicos.

Los expertos que pusieron a punto los criterios de seguridad ante las radiaciones partieron de dos hechos básicos: (1) la humanidad está expuesta a un fondo natural de radiación cuyo efecto biológico no es perceptible, aunque en principio no exista un umbral para los efectos de las radiaciones ionizantes; (2) el daño biológico es proporcional a la energía absorbida por el sistema viviente tomado en consideración, y esta relación lineal es más o menos independiente de la fuente de radiación en concreto. Los expertos han admitido que, en lo que se refiere a las fuentes bien conocidas, las dosis absorbidas podían sobrepasar en varias veces a la dosis derivada de las radiaciones naturales de fondo, y que el riesgo era despreciable en tanto que fuesen inferiores a la dosis necesaria para doblar la tasa de mutación en una población de mamíferos. En general, se adoptaron tales dosis como las aceptadas para la exposición humana. En el caso de nuevos materiales radiactivos, fabricados o descubiertos, de los que no hubiera datos experimentales recopilados, los valores permitidos se calcularían en base a la relación lineal entre el daño biológico y la energía liberada en el interior del organismo.

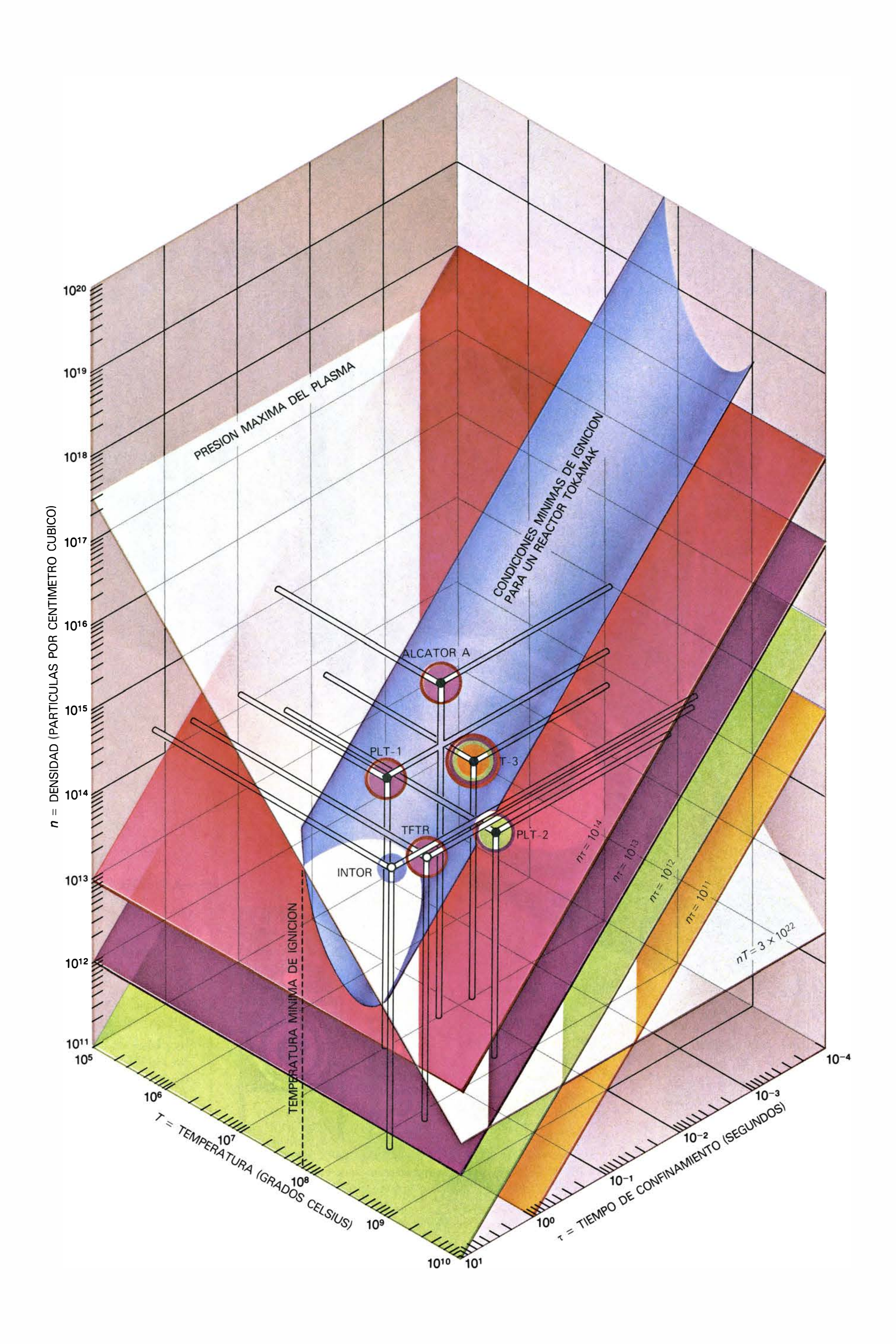
Dado que la polución química es de extensión mundial, un comité internacional de expertos debería proponer normas protectoras para que fueran adoptadas por los distintos gobiernos. Algunos principios lograrían fácilmente una aceptación general: por ejemplo, deberían prohibirse todos los carcinógenos potenciales innecesarios. Debería confeccionarse también una relación de carcinógenos de los que no se puede prescindir, pero la exposición a los cuales debería limitarse estrictamente. Será difícil, sin embargo, alcanzar un consenso sobre una lista de tal tipo y sobre las concentraciones ambientales permisibles de cada compuesto. Uno quisiera comenzar definiendo las concentraciones permisibles de los carcinógenos indispensables en relación con el fondo "natural". El problema es que las diferencias culturales, los hábitos sociales y otros factores dan lugar a ideas muy tergiversadas de lo que es un nivel de fondo natural, de riesgo despreciable.

Si se alcanza un acuerdo sobre el nivel de riesgo despreciable, el siguiente paso sería estimar, para cada carcinógeno indispensable, las concentraciones permitidas a las que pueden estar expuestos los obreros en las fábricas y, a un nivel muy inferior, las que afectan a toda la población. ¿Pueden ayudar los ensayos bacterianos a fijar estos niveles? Sí, si hay proporcionalidad entre la capacidad de dañar el ADN de las bacterias y la capacidad de provocar el cáncer en los mamíferos. Matthew S. Meselson y Kenneth Russell, de la Universidad de Harvard, han intentado demostrar la existencia de una relación directa entre la capacidad mutágena en el test de *Salmonella* y la capacidad cancerígena en animales de laboratorio. En 10 de los 14 carcinógenos químicos conocidos consideraron que había correlación.

Era de esperar la existencia de excepciones a la regla de la correlación ya que en los mamíferos la magnitud de la lesión del ADN en una célula afectada depende de factores (que pueden diferir de los bacterianos) tales como el metabolismo específico del carcinógeno en el órgano afectado, la permeabilidad celular a determinados metabolitos, la tendencia a acumularlos y la dimensión del deterioro del ADN después de haber actuado los mecanismos celulares de reparación. Debe determinarse cada una de estas variables si se pretende precisar niveles de exposición válidos para los seres humanos. Cada compuesto químico, a diferencia de las distintas fuentes de radiación, requiere un completo estudio específico para determinar su potencialidad cancerígena. Los ensayos bacterianos dan valores indicativos aceptables.

Aun cuando establecer normas para los carcinógenos químicos sea mucho más difícil de lo que lo fue para las radiaciones, las leyes deberían dictarse pronto. Es más inteligente y seguro tener normas de seguridad que no tenerlas. Gran parte de la población humana padece cáncer; una persona de cada cinco muere de ello. Si se procura elucidar el mecanismo por el que cada uno de los productos cancerígenos provoca lesiones en el ADN, es probable que se mejore no sólo la prevención del cáncer sino también los medios para curarlo. Alrededor de la mitad de los fármacos administrados para intentar detener el cáncer son agentes que dañan el ADN, con lo que un número significativo de personas quedan expuestas, regularmente y con fines terapéuticos, a dichos agentes. Determinando, mediante estudios con bacterias, el mecanismo por el cual los agentes antitumorales afectan al ADN, los investigadores pueden ayudar a crear medicamentos eficaces para la curación del cáncer sin, a la vez, provocarlo.

Los ensayos bacterianos de carcinógenos potenciales ayudarían a responder a dos preguntas que se hacen la industria, el gobierno y la gente en general. ¿Qué productos químicos son agentes que dañan el ADN? ¿Qué cantidad de cada uno de ellos en nuestro aire, agua, alimento, en los productos que fabricamos y consumimos constituye un riesgo biológico? Para aportar a estas dos cuestiones una respuesta verdaderamente satisfactoria, hará falta plantear una tercera: ¿cuál es el mecanismo por medio del cual cada compuesto químico afecta al ADN? Sólo la investigación fundamental puede dar la respuesta. Ella es la condición para una verdadera prevención del cáncer.



PRESION MAXIMA DEL PLASMA
CONDICIONES MINIMAS DE IGNICION PARA UN REACTOR TOKAMAK
ALCATOR A
PLT-1
T-3
TFTR
PLT-2
INTOR
TEMPERATURA MINIMA DE IGNICION
$n\tau = 10^{14}$
$n\tau = 10^{13}$
$n\tau = 10^{12}$
$n\tau = 10^{11}$
$nT = 3 \times 10^{22}$
n = DENSIDAD (PARTICULAS POR CENTIMETRO CUBICO)
10^{20}
10^{19}
10^{18}
10^{17}
10^{16}
10^{15}
10^{14}
10^{13}
10^{12}
10^{11}
T = TEMPERATURA (GRADOS CELSIUS)
10^{5}
10^{6}
10^{7}
10^{8}
10^{9}
10^{10}
τ = TIEMPO DE CONFINAMIENTO (SEGUNDOS)
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}

Progresos hacia un reactor de fusión tokamak

Se han producido resultados alentadores en el confinamiento de un plasma de hidrógeno supercaliente en una "botella magnética" toroidal. En breve se conocerá la viabilidad del modelo para un reactor de fusión

Harold P. Furth

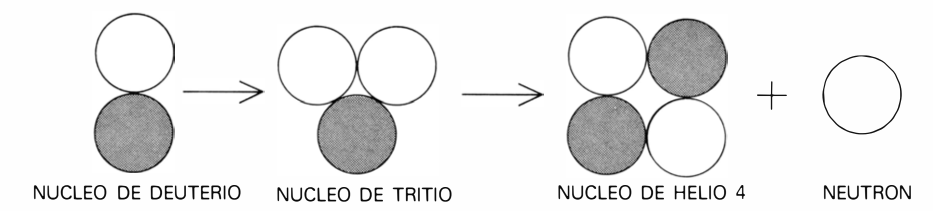
Hará más de un cuarto de siglo que surgió la idea de un reactor de fusión que generase energía, basado en el confinamiento magnético de un plasma (un gas caliente de partículas cargadas) en el interior de una cámara de vacío toroidal. La idea, que no ha perdido vigor, se ha visto reforzada por un número notable de duros experimentos. Si echamos la vista atrás, resulta inmediato que se subestimaron grandemente las dificultades de tipo experimental, si bien desde un comienzo se percibieron cuáles eran los aspectos teóricos fundamentales de la cuestión. Hacia mediados de los años 60, tras 15 años de intensiva investigación en varios países, no había todavía ninguna base experimental en firme para la extrapolación de un modelo de confinamiento magnético a las condiciones del plasma que se consideraban necesarias para un reactor de fusión. Antes bien, había motivos razonables para pensar que, a lo largo de nuestro siglo, no se llegaría a obtener en el laboratorio un rendimiento neto de energía a partir de las reacciones de fusión controladas.

En las postrimerías de la última década, las perspectivas de la investigación sobre la fusión a corto plazo mejoraron notablemente. Los físicos del Instituto de Energía Atómica I. V. Kurchatov, de Moscú, lograron, con éxito, impresionantes resultados en el campo del confinamiento toroidal magnético con un aparato, de relativa sencillez, que llamaron tokamak. En el tokamak ruso Modelo T-3 se obtuvieron, por primera vez, temperaturas del plasma dentro del orden de magnitud de las requeridas para un reactor de fusión toroidal, junto con propiedades de confinamiento plasmático que resultaron ser algo misteriosas pero que, al menos, lo eran favorablemente. Debido a las mejoras conseguidas, aumentaron de forma significativa en todo el mundo los trabajos de investigación sobre la fusión. El esfuerzo se centró, sobre todo, en el estudio del modelo tokamak.

A lo largo de estos últimos diez años, la investigación en el campo de distintos tipos de tokamak, acometida en la Unión Soviética, los Estados Unidos, Europa occidental y Japón, sobre todo, ha ido refinando firmemente los parámetros de funcionamiento de estos aparatos de experimentación. Seguían subsistiendo, empero, graves interrogantes sobre las propiedades de confinamiento que podían esperarse bajo condiciones similares a las que ocurren en un reactor de plasma real, punto éste que acaba de someterse a una prueba crucial al hallarse las condiciones de plasma próximas a las requeridas por un reactor de fusión en un experimento realizado en un tokamak a gran escala en el Laboratorio de Física de Plasmas de la Universidad de Princeton, experimento bautizado con la expresión "Gran Toro de Princeton", PLT (del inglés, "Princeton Large Torus"). Para sorpresa y alegría de los experimentadores, no se observó ningún fallo en la calidad del confinamiento.

El confinamiento del plasma no es el único extremo científico que necesita resolverse antes de la construcción de un reactor de fusión tokamak que sea rentable. Pero en torno a ello podemos afirmar, por primera vez en la historia de la investigación en el campo de la fusión, que se dispone ya de una buena base experimental para una descripción física pormenorizada de dicho reactor. La demostración real de que un reactor de este tipo sea viable científicamente aguarda a que concluya, en los primeros años de la década próxima, la nueva generación de enormes instalaciones de tokamaks. Tres de éstos se hallan en fase

HITOS HISTORICOS que jalonan el camino hacia la obtención de energía por un reactor de fusión práctico, basado en el enfoque tokamak. Se han dibujado, en la página opuesta, con ayuda de una gráfica tridimensional. Los puntos negros simbolizan los mejores funcionamientos, hasta la fecha, de tokamaks experimentales, representados en función de los tres parámetros fundamentales de la investigación en el terreno de los plasmas sometidos a confinamiento magnético, a saber: temperatura (T), densidad (n) y tiempo de confinamiento (τ). El primer éxito sobresaliente, recogido aquí con su notación T-3, se obtuvo en 1969 en un experimento llevado a cabo en el tokamak Modelo T-3, en el Instituto de Energía Atómica I. V. Kurchatov, de Moscú. El valor correspondiente al Alcator, el más importante de una serie de experimentos realizados durante los últimos años en el Instituto de Tecnología de Massachusetts, indica las mejores condiciones de confinamiento conseguidas hasta la fecha en un aparato tokamak que utiliza sólo calefacción óhmica o resistiva del plasma. Aparece dos veces el logro más feliz de operatividad que alcanzó el Gran Toro de Princeton, o PLT, un tokamak a escala intermedia que entró en funcionamiento en 1976 en el Laboratorio de Física de Plasmas de la Universidad de Princeton; la primera vez se cita por una prueba sobre calefacción solamente óhmica (PLT-1) y, la segunda, por una prueba sobre calefacción óhmica más calefacción auxiliar de haz neutro (PLT-2). Los puntos blancos señalan los cálculos de las características de funcionamiento esperadas para dos próximos aparatos experimentales, uno de los cuales, el Reactor de Fusión de Prueba Tokamak (TFTR), se halla en fase de construcción en Princeton. El otro punto blanco representa el funcionamiento previsto para el Reactor Internacional Tokamak (INTOR), ingenio de mayores proporciones que se encuentra en fase de proyecto. El conjunto de planos paralelos en color designa incrementos sucesivos de orden de magnitud del parámetro de funcionamiento, que es una medida de la calidad de funcionamiento del plasma, y que se expresa como el producto de la densidad de partículas por centímetro cuadrado por el tiempo de confinamiento dado en segundos ($n\tau$). La superficie azul, curvada, del extremo superior del conjunto es una gráfica del valor mínimo de $n\tau$, que hay que superar para conseguir la ignición o autocalefacción, régimen de un plasma que consta de deuterio y tritio (los dos isótopos pesados del hidrógeno). El plano blanco que atraviesa todas las superficies es la presión máxima del plasma (nT) que puede mantenerse en un reactor tokamak. Se espera que el reactor internacional INTOR funcione entre el plano blanco y la superficie azul curvada, y sea el primer reactor tokamak de ignición.



REACCION DE FUSION entre un núcleo de deuterio y un núcleo de tritio. Esta reacción da lugar a un núcleo de helio (una partícula alfa), un neutrón y 17,6 millones de electronvolt (17,6 MeV). Para alcanzar esa fusión, los dos núcleos que entran en choque deberán moverse con una energía cinética relativa, por lo menos, de 10.000 electronvolt (10 keV). Un plasma que conste de partículas cargadas que se desplacen con una energía media aleatoria de 15 keV por partícula poseerá una temperatura de 10 keV, unos 116 millones de grados C. Las bolas grises son protones y, neutrones, las blancas.

de construcción: uno en los Estados Unidos, otro en Europa y el tercero en Japón. Dado ese paso crítico, habrá que despejar algunas incógnitas fundamentales en el terreno de su tecnología para que sea un hecho el funcionamiento de un reactor de fusión que genere energía.

Para obtener, en el laboratorio, la reacción de fusión más sencilla se juntan los núcleos de dos isótopos pesados del hidrógeno: el deuterio y el tritio. Ello se logra tras vencer la repulsión electrostática de Coulomb entre esos dos núcleos cargados positivamente antes de que la fusión de ambos de origen a un núcleo de helio. Por ello, hay que aumentar al comienzo la energía cinética relativa (energía de movimiento) de los núcleos. En el caso de una reacción de fusión con deuterio y tritio, la inversión mínima de energía es de unos 10.000 electronvolt (10 keV). Si lo miramos con la óptica del físico teórico que trata la energética global del proceso de fusión, esta inversión es bastante modesta, pues la energía liberada en cada suceso de fusión es de 17,6 millones de electronvolt (17,6 MeV). Mas, para el físico experimentador, las exigencias son de un calibre altísimo: una energía media aleatoria de 15 keV por núcleo equivale a una temperatura del combustible de 10 keV, que traducido en grados Celsius suma la cifra de 116 millones.

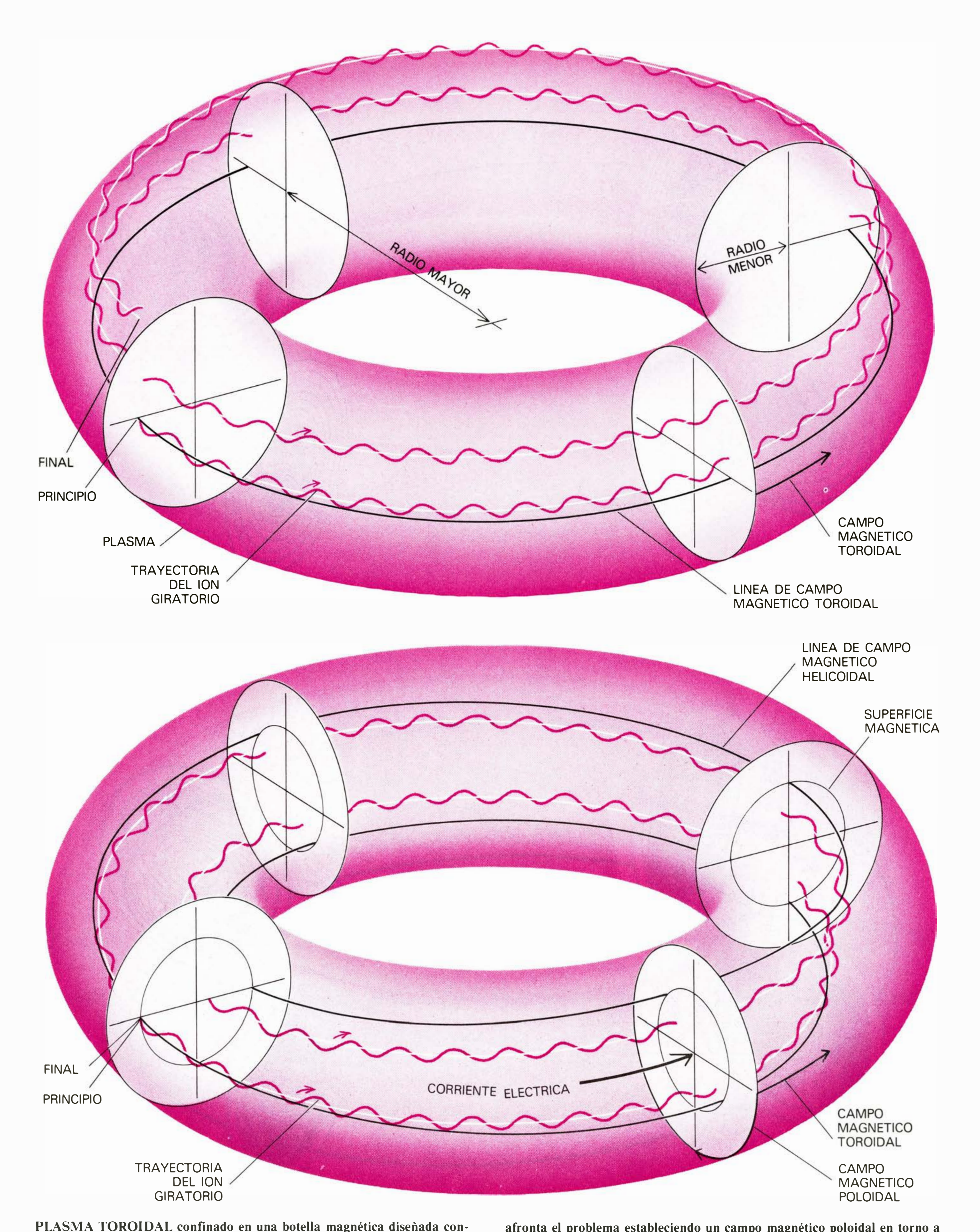
Que la materia se confine a temperaturas tan elevadas se explica por la disociación de los átomos en sus constituyentes cargados eléctricamente –los electrones y los iones– que tiene lugar a temperaturas superiores a los 10.000 grados Celsius. Las partículas cargadas tienden a girar a lo largo de las líneas de un campo magnético; por esa razón, podemos confinar el plasma resultante en una "botella magnética" de un modo adecuado. La botella magnética toroidal goza de la ventaja especial de que las partículas cargadas no pueden escaparse en virtud de un simple desplazamiento a lo largo de las líneas de fuerza del campo magnético. La máquina espejo, dispositivo que consiste en una botella de extremo abierto, constituye la principal vía alternativa del reactor de fusión de las que se basan en el confinamiento magnético de un plasma caliente. Existen, por otro lado, varios modelos de "confinamiento inercial" completamente distintos, en los que haces intensos de luz láser o de partículas aceleradas implotan una pequeña bolita de combustible; el combustible se expande luego libremente y se fusionan sus núcleos de hidrógeno pesado.

En un reactor de fusión experimental, la cifra más importante a averiguar es la que viene dada por la razón entre la potencia de salida generada por las reacciones de fusión y la potencia de entrada que se necesita para calentar el plasma del reactor. Esa razón, que se denomina factor de multiplicación de la energía, o valor Q, depende de la fracción de los núcleos calientes que pueden fusionarse durante el tiempo que el plasma tardará en perder su energía. Al ser la fusión un proceso binario, el valor Q depende del producto del tiempo de confinamiento de energía del plasma (expresado en segundos) por la densidad del plasma (medida en partículas por centímetro cúbico). Por ejemplo, en un plasma de deuterio-tritio a 10 keV con una distribución normal de las energías de las partículas, denominada "requisito de igualación" (energía entrante igual a energía saliente, lo que corresponde a un valor de Q igual a la unidad) significa que el producto del tiempo de confinamiento por la densidad, una medida conocida como el parámetro del confinamiento, debería exceder de 6×10^{13}. El valor Q de un plasma aumenta rápidamente una vez que el parámetro de confinamiento supera el umbral de igualación, ya que un 20 por ciento de la energía liberada en las reacciones de fusión en las que interviene deuterio y tritio entra en la producción de núcleos de helio-4 energéticos (partículas alfa), que pueden frenarse y retenerse en el seno del plasma, con lo que aumenta la energía de entrada disponible para calefacción. Cuando el parámetro de confinamiento se aproxima a 3×10^{14}, las reacciones de fusión posibilitan el mantenimiento de la temperatura del plasma, sin necesidad de recurrir a ningún aporte externo de energía calorífica. En ese instante, el valor Q se hace infinito y se dice que el combustible "arde" en el sentido convencional.

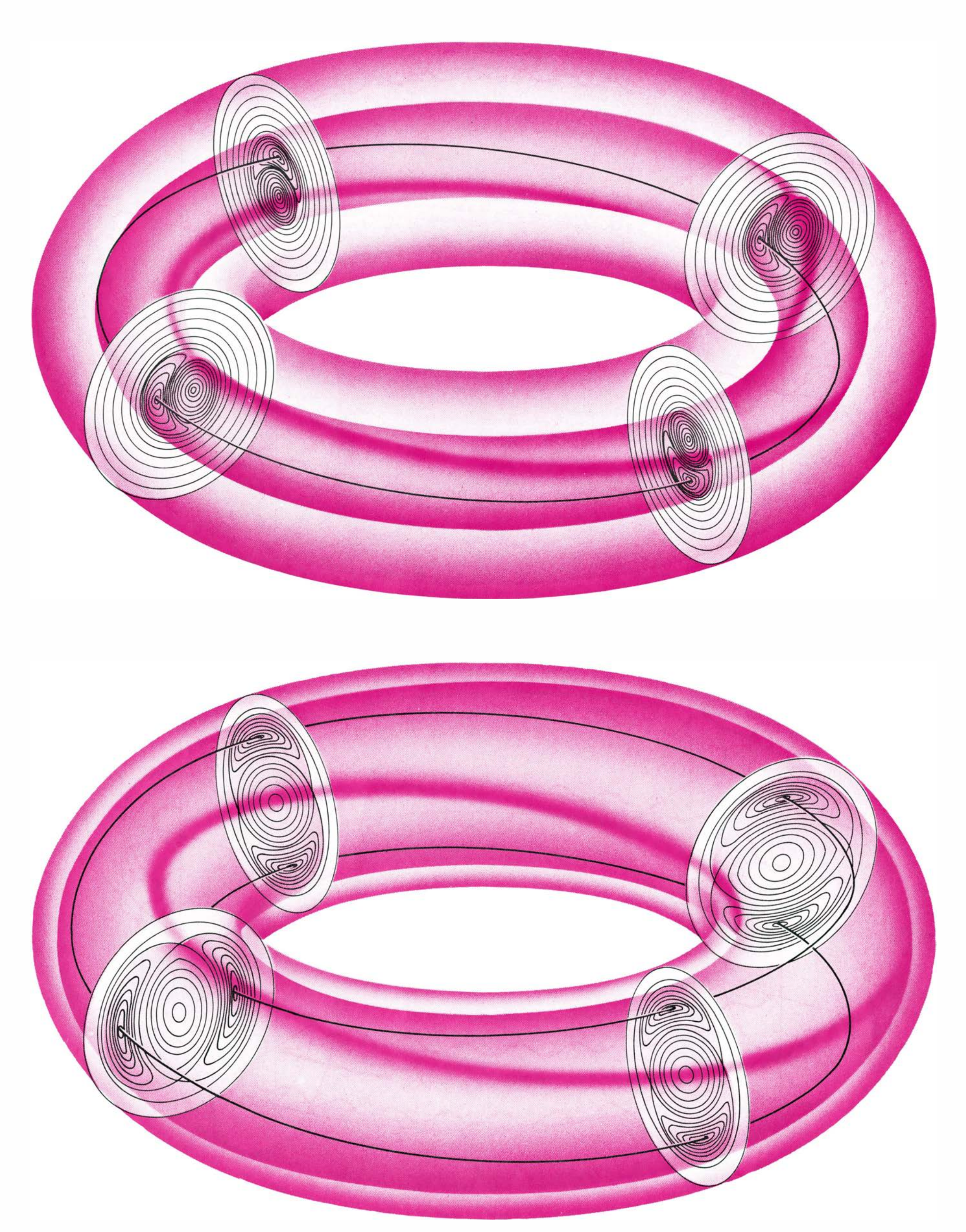
Los investigadores en el campo de la fusión, en su afán por conseguir el grado de confinamiento que precisa un reactor, han llegado a delimitar cuatro procesos principales de pérdida de energía que deben mantenerse sometidos a control: (1) inestabilidades macroscópicas, es decir, deformaciones magnetohidrodinámicas espontáneas del campo confinante que provoca pérdidas notables en el plasma a través de la botella magnética; (2) inestabilidades microscópicas, que son leves pérdidas paulatinas que permiten que las partículas del plasma (y el calor) se difundan gradualmente fuera del plasma a través de las líneas de campo magnético; (3) difusión de Coulomb, que rompe las órbitas de las partículas que giran y da lugar a una razón de difusión irreducible; y (4) enfriamiento radiativo del plasma, sobre todo en la forma de radiación ultravioleta emitida por los iones de impureza. La comprensión creciente de esos cuatro procesos de pérdida de energía y el desarrollo de técnicas eficaces para minimizarlas han traído como consecuencia los progresos realizados en la investigación de aparatos toroidales de fusión.

Si debemos confinar un plasma con la forma de toro, la botella magnética que lo confine no puede constar sólo de un campo magnético toroidal; en ese caso, las órbitas de las partículas no se cerrarían sobre ellas mismas, y las partículas escaparían con facilidad de la botella a través de las líneas de campo magnético.

El problema puede abordarse desde distintos enfoques, siendo uno la llamada aproximación del estellator, modelo toroidal que desarrolló originalmente Lyman Spitzer, Jr., de Princeton. Pero la solución más simple, y con mucho, la proporciona el tokamak. En este modelo, propuesto de modo independiente en los comienzos de la década de los 50 por Igor E. Tamm y Andrei D. Sakharov, de Moscú, y por Spitzer, de Princeton, se induce una corriente eléctrica a lo largo de la columna de plasma, que fluye alrededor del toro. El campo magnético poloidal asociado (cuyas líneas de fuerza giran en torno a la sección transversal menor del toro) origina líneas de campo helicoidales que forman un conjunto de superficies magnéticas en-



PLASMA TOROIDAL confinado en una botella magnética diseñada convenientemente. El plasma toroidal muestra la ventaja especial de que las partículas cargadas no pueden escaparse del mismo porque se desplazan a lo largo de las líneas del campo magnético. Pero una configuración magnética de este tipo no puede consistir sólo en un campo magnético toroidal, toda vez que, entonces, no se cerrarían las órbitas de las partículas y estas últimas acabarían por escaparse del plasma moviéndose a través de las líneas del campo magnético (*ilustración superior*). El enfoque tokamak afronta el problema estableciendo un campo magnético poloidal en torno a la sección transversal del toro plasmático mediante una corriente eléctrica inducida a lo largo de la columna del plasma. Al combinarse el campo anterior con un campo magnético toroidal aplicado externamente, se originan líneas de campo helicoidales; éstas forman un conjunto de superficies magnéticas encajadas (*ilustración inferior*). Una partícula que en la última configuración se desplace alrededor del toro no se desviará apreciablemente de una superficie magnética dada. (Ilustraciones de George V. Kelvin.)



DOS BUCLES INDESEABLES que pueden aparecer de repente en un plasma tokamak a altas densidades de corriente, y amenazar así la estabilidad macroscópica del plasma. Puesto que estas deformaciones magnetohidrodinámicas se hallan alineadas, con bastante aproximación, con el paso de rosca de las líneas de campo magnético helicoidales, existe una relación sencilla entre la helicidad de las líneas de campo y la forma de los modos en bucle característicos del plasma. El diagrama de arriba muestra un bucle lateral que gira una vez alrededor del toro en la dirección toroidal cada vez que da una vuelta a su alrededor en dirección poloidal. El diagrama inferior simboliza un bucle elíptico, que va dos veces en torno al toro en dirección toroidal por una vuelta en dirección poloidal. Caben modos de inestabilidad por flexión de órdenes superiores, pero el plasma tiende a ser más resistente a tales deformaciones complejas. El bucle lateral desempeña un papel importante en el mantenimiento del canal de corriente en el centro del plasma.

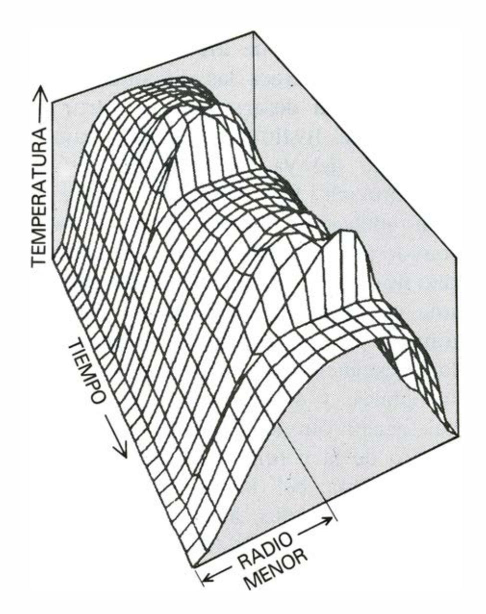
cajadas. En esa configuración, las partículas cargadas tienen órbitas cerradas que no se separan de una superficie magnética dada en más de sus radios de giro, medido en la dirección del campo poloidal.

Esta localización de las órbitas de las partículas es importante a la hora de definir la razón irreducible de difusión de partículas que resulta de la difusión de Coulomb, que puede causar que las partículas salten la anchura orbital desde una superficie magnética a otra. Para iones instalados en el rango de los 10 keV, la probabilidad de que dos núcleos se fusionen en vez de difundirse tan sólo es muy pequeña (alrededor de una parte en diez millones). Para minimizar la importancia de este proceso de difusión de calor particular que dificulta la obtención de potencia de fusión neta, el radio menor del toro de plasma debe ser, por consiguiente, varias veces mayor que el diámetro orbital de las partículas. Al reducir el diámetro orbital cuando aumenta la intensidad del campo poloidal, el criterio para la ignición (la obtención de un parámetro de confinamiento mayor que 3×10^{14}) en un tokamak resulta ser aproximadamente equivalente a la condición de que la corriente toroidal supere los cuatro millones de ampère. Cuando ello sucede, se necesita esa misma corriente mínima de unos cuatro millones de ampère para localizar las órbitas de las partículas alfa energéticas en la columna del plasma, condición previa para lograr la ignición.

Aunque en un tokamak el confinamiento microscópico de las partículas plasmáticas mejora al elevarse la corriente del plasma, la estabilidad macroscópica de éste adquiere una enorme inseguridad. El plasma puede deformarse de repente en varias configuraciones más o menos retorcidas, llamadas en bucle, que nos recuerdan los modos de alabeado mecánico de las componentes de una estructura sometida a una tensión excesiva. Los bucles inestables del plasma del tokamak tienden a alinearse, de una manera bastante aproximada, con el paso de rosca de las líneas helicoidales de campo magnético. Resulta así una relación sencilla entre el "factor de seguridad" (una medida de la helicidad de las líneas de campo) y la forma de los modos en bucle característicos del plasma. Asignamos, por ejemplo un factor de seguridad igual a la unidad para el bucle que va una vez alrededor del toro en la dirección toroidal por cada vez que lo rodea en dirección poloidal, lo que corresponde a un simple desplazamiento lateral del plasma en la sección transversal menor del toro (primer modo); un factor de seguridad igual a dos, si el bucle da dos vueltas alrededor del toro en dirección toroidal por cada una que lo hace en dirección poloidal, lo que corresponde a una deformación elíptica (segundo modo); un factor de seguridad igual a tres, si el bucle va tres veces alrededor del toro en dirección toroidal por cada vuelta al mismo en dirección poloidal, lo que corresponde a una deformación triangular (tercer modo), y así sucesivamente.

Para las corrientes típicas del tokamak, podemos demostrar teóricamente que cuanto mayor sea el factor de seguridad (es decir, cuanto menos profundo sea el paso de rosca de las líneas de campo helicoidales), más alto será el número del modo de las posibles inestabilidades y más resistente a la deformación será el plasma. Por consiguiente, cuanto menor sea la corriente del plasma, mayor resultará la estabilidad macroscópica de éste. Se ha observado que el comportamiento experimental de la columna plasmática de distintos tokamaks es coherente con estas predicciones, y con otras más detalladas, de la teoría magnetohidrodinámica.

En resumen, el fin práctico que se propone la investigación en el terreno de los tokamaks es lograr que la helicidad de las líneas de campo sea lo más empinada posible, al tiempo que se mantiene una buena estabilidad macroscópica; y ello, para obtener la máxima corriente plasmática que posibilite el mejor confinamiento de las órbitas de las partículas. Se impone, pues, dominar el manejo de los dos modos en bucle que ostentan los números modales más bajos: la deformación lateral (primero) y la deformación elíptica (segundo). No es fácil proyectar teóricamente una configuración de campo que sea estable con un factor de seguridad adecuadamente bajo. A pesar de lo cual, el plasma del tokamak llega automáticamente a una solución de trabajo. Puesto que la parte más caliente del canal de corriente es la más conductora, la densidad de corriente se eleva con lentitud en las proximidades del eje magnético del toro hasta que el valor local del factor de seguridad cae por debajo de la unidad; como resultado de ello, surge una inestabilidad lateral y se amplía el canal de corriente nuevamente. Ese mecanismo de reajuste se localiza en la región central del plasma y no causa, por necesidad, una pérdida neta de plasma caliente en la pared de la cámara de vacío. La existencia de una región central con un factor de seguridad inferior a la unidad ofrece la gran ventaja de que, para este tipo de configuración de campo magnético, no puede desarrollarse en absoluto el modo en bucle elíptico, más inoportuno y que se sitúa hacia la periferia, aun cuando el factor de seguridad valga dos para radios mayores. El resultado es eficazmente estable en todo el confinamiento plasmático. Y es ese mecanismo feliz de autocontrol que posee el plasma del tokamak el responsable, en un sentido darwinista, de la aparición del tokamak como la botella magnética toroidal más viable.



OBSERVACIONES con rayos X de un bucle real de inestabilidad, realizadas en un plasma tokamak. Dichas observaciones se aprovecharon para la construcción, por ordenador, de la gráfica tridimensional superior. Los perfiles de temperatura oscilan a medida que van creciendo hasta un pico, para bajar luego rápidamente. Los datos de rayos X se obtuvieron durante una prueba de funcionamiento del aparato PLT. Ned R. Sauthoff verificó el proceso de los resultados experimentales.

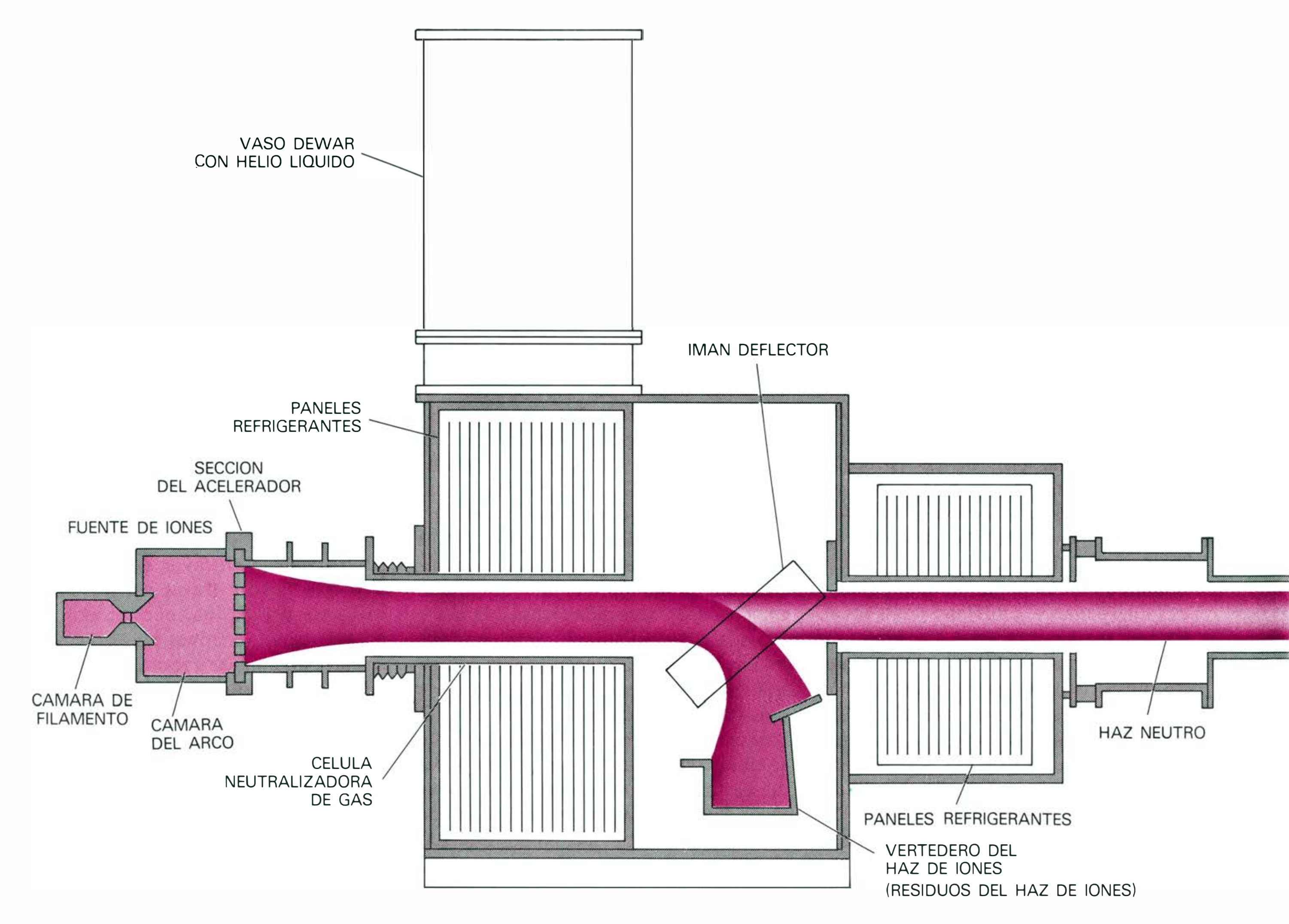
Otra ventaja en favor de la corriente que fluye a través del plasma del tokamak estriba en que proporciona energía a los electrones plasmáticos por calefacción óhmica o resistiva. En virtud de ello, se obtienen espontáneamente en las descargas del tokamak temperaturas electrónicas del orden del kilovolt. Los electrones que giran chocan con los iones; con el tiempo, alcanzan un estado de equilibrio térmico con ellos hasta un punto que queda determinado por la temperatura electrónica y el parámetro de confinamiento. En los plasmas con un parámetro de confinamiento pequeño, la temperatura de los iones es típicamente mucho más baja que la de los electrones, pero a medida que dicho parámetro se va acercando al umbral de igualación, las temperaturas de unos y otros tienden a igualarse.

En el experimento T-3 se consiguió, mediante calefacción óhmica, que el

mayor tokamak de los existentes hace diez años superara las dificultades. (El T-3 se había desarrollado, a partir de 1962, en el Instituto Kurchatov bajo la dirección del ya desaparecido Lev A. Artsimovich.) La máquina T-3 posee un radio menor de 15 centímetros, un radio mayor de 100 centímetros, una intensidad de campo toroidal de 35.000 gauss y una corriente característica de 100.000 ampère. Por vez primera en el terreno de la investigación sobre aparatos de fusión toroidales, y siguiendo un proceso de calefacción óhmica, la temperatura en el centro de la columna plasmática subió por encima del kiloelectronvolt y la temperatura iónica llegó a los 0,5 keV, aproximadamente. El tiempo de confinamiento de energía se situó entre cinco y 10 milisegundos, con una densidad central que rondaba los 10^{14} de partículas por centímetro cúbico. Por tanto, el parámetro de confinamiento para el T-3 estuvo cerca de 10^{12}, valor todavía alejado del que debería conseguir un reactor, aunque constituía un guarismo mucho más alto que el logrado hasta entonces en cualquier otro experimento de fusión.

Antes del experimento T-3 se habían obtenido parámetros de confinamiento característicos para máquinas toroidales del orden de 10^{11}, a temperaturas electrónicas de unos 100 electronvolt. Es más, los tiempos de confinamiento observados no habían tenido ninguna relación con el ventajoso tipo de confinamiento que se pretendía aportase la geometría toroidal; por el contrario, parecía como si siguieran leyes de escala anómalas, incompatibles con la consecución de condiciones de fusión en un reactor no excesivamente grande. La presencia de varias microinestabilidades, que en algunos casos podían observarse directamente, era la razón que se sospechaba como la causante del pobre confinamiento del plasma que empeoraba al elevarse la temperatura. El único elemento alentador en esta situación provenía de la teoría predominante sobre microinestabilidad, que predecía que tan grandes proporciones de transporte anómalo de plasma no se mantendrían a temperaturas aún más altas.

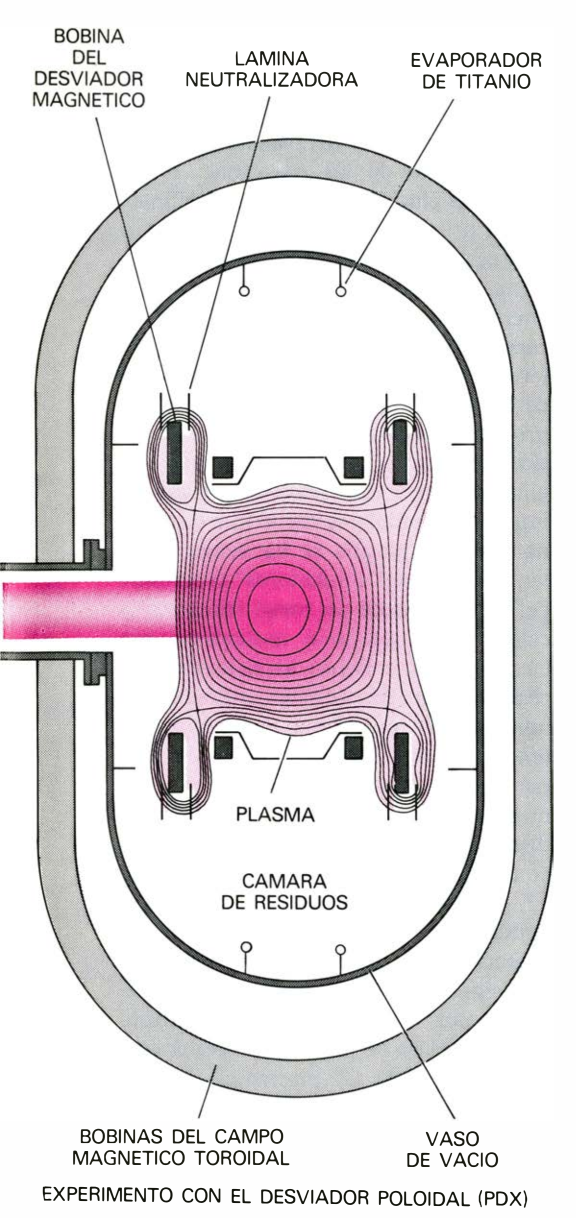
Los nuevos datos del experimento T-3 justificaban esta expectación. Por vez primera en tales experimentos se conseguían grandes tiempos de confinamiento, que se aproximaban razonablemente a los valores predichos en la teoría de la difusión coulombiana. Los iones del plasma parecían seguir, muy de



DOS AVANCES IMPORTANTES en el diseño de experimentos de fusión con el tokamak. A la izquierda, el aparato que aparece en sección transversal es un acelerador de haz neutro, una fuente de calefacción auxiliar del plasma desarrollada básicamente en el Laboratorio Nacional de Oak Ridge y el Laboratorio Lawrence de Berkeley. Se inyecta en el tokamak el haz energético de átomos neutros de hidrógeno producidos por el acelerador; pasan libremente a través de las líneas de campo magnético; al chocar con las partículas plasmáticas, pierden sus electrones y son retenidos en el plasma como iones energéticos cargados positivamente. A la derecha, un tokamak en sección transversal equipado con un desviador magnético, un ajuste en el que se emplea un separador de campo magnético que evita que penetren iones de impureza en la columna principal del plasma y se reduce, por tanto, la tasa de calor perdida por radiación. Los iones de impureza se pegan sobre una película de titanio aplicada sobre las superficies interiores de las dos "cámaras de residuos", que operan a modo de bomba de vacío. La ilustración recoge el Experimento del Desviador Poloidal; éste puede funcionar de varios modos distintos, originando la sección transversal del plasma, en cuadrado, que se muestra, o en forma de *D* (lineal o invertida).

cerca, las predicciones teóricas; se hallaba que sólo los electrones se acomodaban, en una forma residual de transporte anómalo, con leyes de escala que permanecían algo dudosas aunque parecían compatibles con un reactor de fusión razonablemente pequeño.

Se precisaban dos mejoras más para progresar, a partir de los resultados del T-3, hasta las condiciones de un reactor real. Debía aumentarse el parámetro de confinamiento en unos dos órdenes de magnitud más y la temperatura tenía que elevarse en un orden de magnitud. La forma más obvia de mejorar el confinamiento era recurrir a un plasma mayor, porque, para procesos difusivos de pérdida de calor, el tiempo de confinamiento de energía aumenta con el cuadrado del radio menor del plasma. El plasma característico de un reactor de fusión tokamak posee un radio menor de uno o dos metros, que supera en unas 10 veces el radio menor del plasma del T-3. Por consiguiente, se espera que el tiempo de confinamiento en un reactor tokamak sea 100 veces mayor que el del experimento T-3, con tal que la razón de difusión no aumente a temperaturas más elevadas. Una vía alternativa para mejorar el confinamiento sería aumentar la intensidad del campo magnético; pero aquí las opciones son más limitadas, pues quizá la máxima intensidad de campo compatible con las bobinas magnéticas superconductoras no rebase los 70.000 gauss. En cuanto al factor 10 que le falta a la temperatura del plasma, la respuesta obvia es aumentar el suministro de potencia calefactora del plasma por calefacción óhmica adicional o por algún otro método auxiliar.



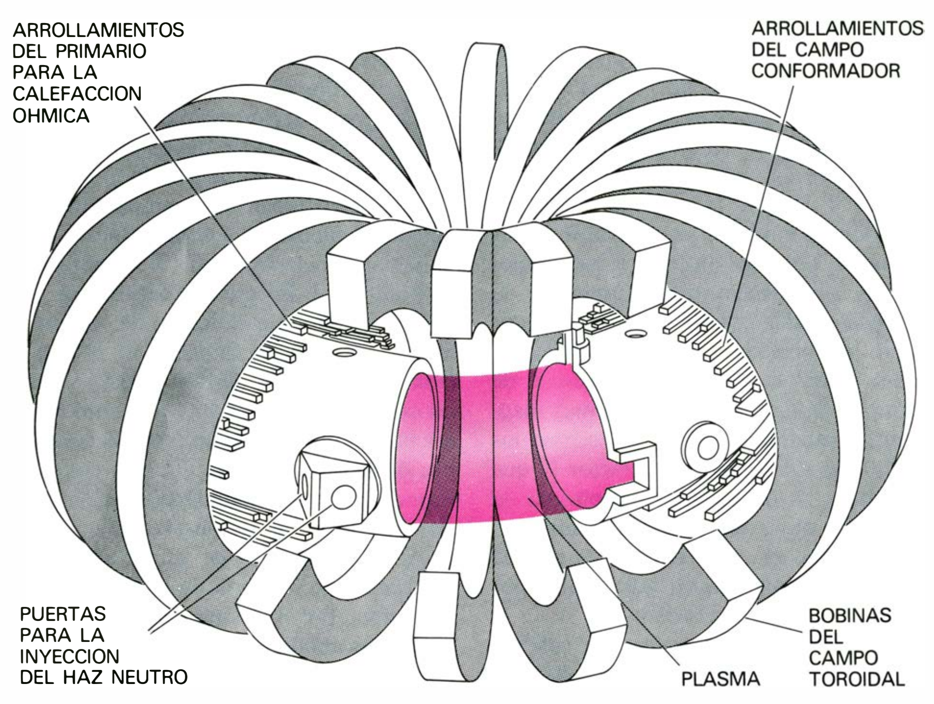
EXPERIMENTO CON EL DESVIADOR POLOIDAL (PDX)

Para los fines del reactor, la calefacción exclusivamente óhmica resulta problemática. En efecto, la resistividad del plasma decrece como función inversa de la temperatura electrónica y la densidad de corriente decrece inversamente al aumentar el tamaño del plasma. Al acercarse a las condiciones del reactor, la densidad de potencia suministrada por calefacción óhmica tiende a disminuir en proporción a los procesos rivales de pérdida de calor (difusión y radiación). En un plasma de deuterio-tritio, vendría a salvar la situación en último término el calentamiento de las partículas alfa del plasma, si la calefacción óhmica pudiera elevar la temperatura del plasma por encima de los cuatro o cinco keV. Esta posibilidad se ha examinado exhaustivamente, pero la conclusión a que se ha llegado es que ahí nos enfrentamos con un verdadero escollo, en particular si lo que se pretende es un reactor tokamak que genere energía y no un experimento más para conseguir, y estudiar, el proceso de ignición.

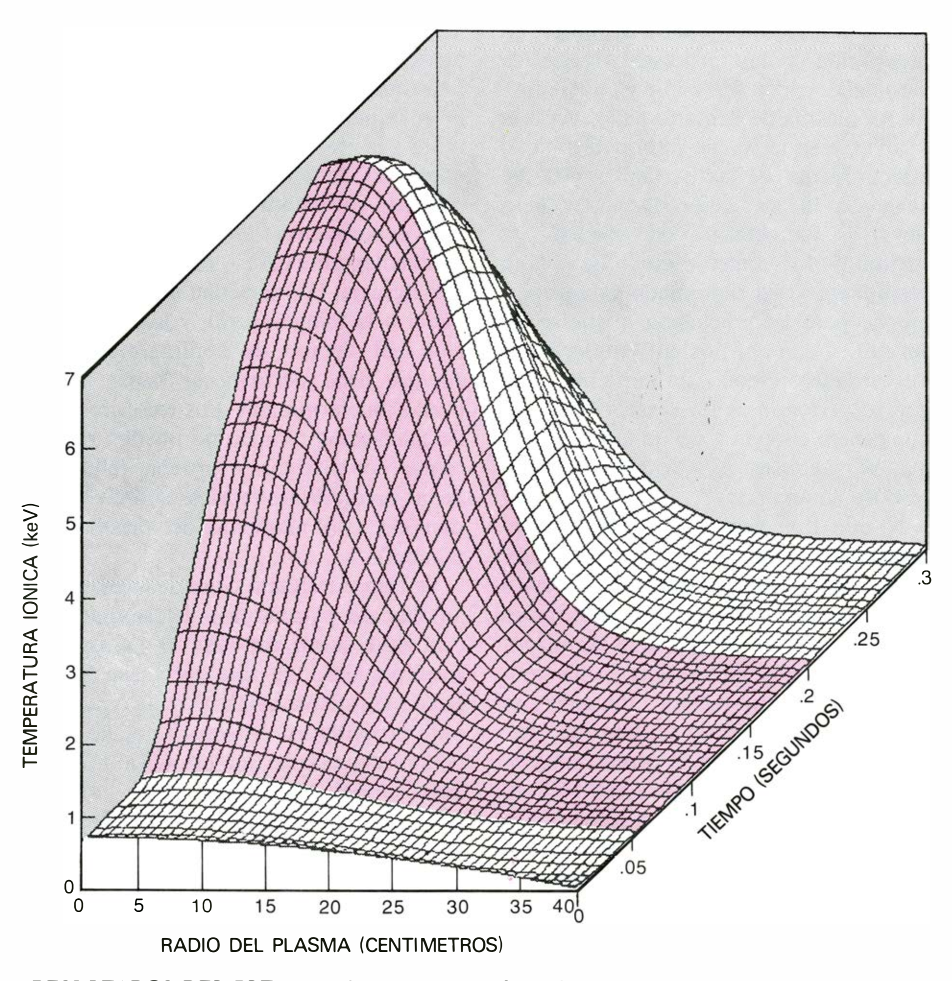
Ni que decir tiene que cualquier experimento con plasmas que roce las condiciones de ignición resulta de un claro interés científico. De las consideraciones señaladas arriba se sigue que las mayores posibilidades de avanzar en las condiciones de un reactor tokamak se hallan en el uso de un plasma muy pequeño conjugado con el campo magnético más intenso. Se aplica esa idea en los aparatos Alcator del Instituto de Tecnología de Massachusetts. El plasma del primero de estos tokamaks, el Alcator *A*, tiene un radio menor de 10 centímetros y un radio mayor de 54 centímetros. En su funcionamiento, se han logrado intensidades del campo magnético de 85.000 gauss, obtenidas con bobinas de cobre de modelo Bitter (llamadas así en honor de Francis Bitter, del MIT). En experimentos hechos con este aparato en los últimos años, la densidad de potencia de calefacción óhmica ha sido suficiente para permitir que plasmas con una densidad de partículas en el centro superior a 10^{15} electrones por centímetro cúbico se calentaran a temperaturas próximas al kiloelectronvolt (11,6 millones de grados C.). En estos experimentos se ha encontrado que el tiempo de confinamiento de energía aumenta en proporción a la densidad de partículas, lo que constituye una muestra de comportamiento anómalo originado por la contribución dominante de la difusión electrónica al proceso de conducción de calor. Por otra parte, se ha observado que los iones se comportan de acuerdo con la predicción de la teoría de la difusión de Coulomb. En el Alcator *A*, se obtuvo, bajo estas condiciones, un parámetro de confinamiento máximo de 3×10^{13}, lo que representaba una mejora substancial sobre los experimentos en tokamaks más antiguos. Recientemente ha entrado en funcionamiento el Alcator *C*, una máquina todavía más potente.

Los experimentos sobre calefacción óhmica realizados en los Alcator y en otros tokamaks de distintos tamaños e intensidades de campo han suministrado gran riqueza de información en lo referente a la elevación del tiempo de confinamiento de energía característico del régimen del tokamak. Sin embargo, en tales experimentos no cabe variar independientemente ciertos parámetros clave (verbigracia, la intensidad de campo poloidal o la temperatura); y así, la elevación del tiempo de confinamiento de energía debe permanecer hasta cierto punto indeterminada. Los experimentos de calefacción óhmica no pueden representar por sí solos la prueba definitiva de las predicciones teóricas sobre el régimen del confinamiento del plasma del tokamak.

Se impone alguna clase de calefacción auxiliar del plasma para determinar con exactitud cómo depende el tiempo de confinamiento de energía respecto de estos parámetros clave. Hay otro motivo para proceder así. Se espera que las partículas del plasma en un reactor tokamak tengan recorridos libres medios muy largos (o pequeñas probabilidades de difusión) debido a su alta temperatura y su densidad comparativamente baja. En este régimen virtualmente sin choques, la teoría predice que una nueva familia de microinestabilidades, llamadas modos de partícula atrapada, pudiera convertirse en el mecanismo dominante de pérdida de energía de un



ESTRUCTURA INTERNA DEL PLT, que hoy constituye uno de los mayores tokamaks en funcionamiento en todo el mundo. Dentro del PLT, el toro plasmático tiene un radio mayor de 130 centímetros y un radio menor de 45 centímetros. Los experimentos que se acometen en él están diseñados para comprobar distintos métodos de calefacción del plasma: calefacción óhmica, calefacción con haz neutro y calefacción con microondas. En el reactor tokamak suelen instalarse varios aparatos de diagnóstico para controlar las propiedades del plasma. (El gran Toro de Princeton que venimos abreviando por PLT, del inglés "Princeton Large Torns", está instalado en el laboratorio de física de plasmas universitario.)



RESULTADOS DEL PLT, en gráfica tridimensional sacada por ordenador, para la temperatura en función del tiempo y la distancia al centro del plasma. Las zonas coloreadas muestran la duración del pulso del haz neutro. El valor máximo de unos 6,8 keV corresponde a una temperatura iónica más de 75 millones de grados Celsius, un nuevo record en la historia de la investigación sobre el tokamak.

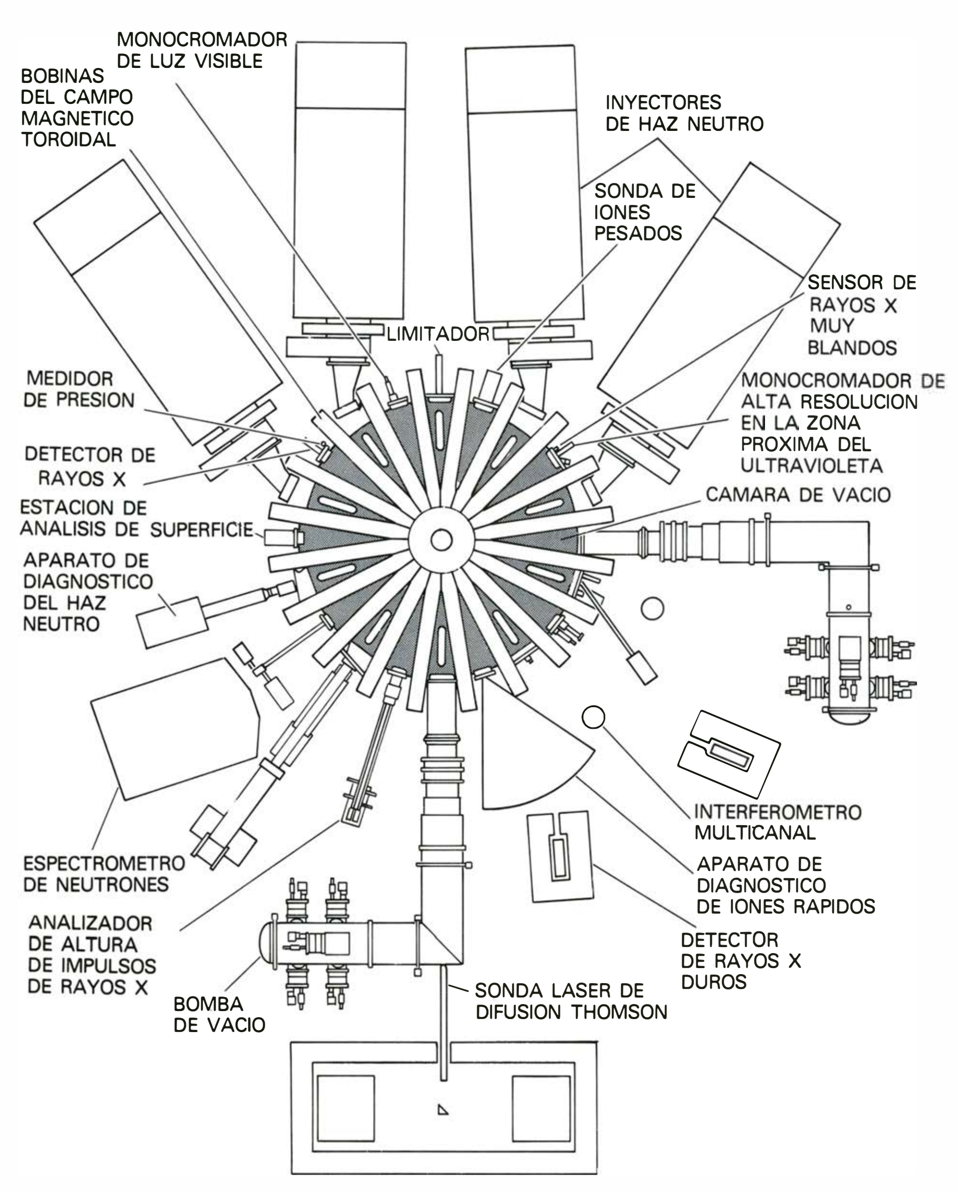
tokamak. Al examinar plasmas con alta proporción de choque de partículas (es decir, plasmas con una temperatura moderada y una densidad alta), los experimentos de calefacción óhmica pudieron conseguir parámetros de confinamiento bastante altos, pero dejaron sin respuesta a la cuestión de cómo se comportaría el régimen sin choques de un reactor. La tarea de examinar la dependencia entre la alta temperatura crítica y el tiempo de confinamiento de energía reclama experimentos realizados en un tokamak con intensas fuentes auxiliares de calefacción. En los primeros años 70, se acometieron varios proyectos que, recientemente, han empezado a producir resultados importantes.

Antes de adelantar las dificultades de la calefacción auxiliar del tokamak es oportuno proponer una cuestión de gran importancia incluso en la fase de la calefacción óhmica: dado el gran flujo de calor que llega al plasma, ¿a dónde va todo el calor cuando refluye fuera del plasma? El transporte radial de calor a través de las superficies magnéticas consiste en un flujo hacia fuera de las partículas más calientes del plasma y un flujo hacia dentro de las más frías. En el borde del plasma, el cual queda definido por la primera superficie magnética que intercepta cualquier estructura sólida externa, hay además un flujo de partículas energéticas a lo largo de las líneas de campo magnético hacia la estructura interceptante. La concentración resultante de partículas de bombardeo tiende a producir el chisporroteo o la propia fusión de la estructura sólida, lo cual origina la afluencia al plasma de iones de impureza. Por ello es por lo que tradicionalmente el plasma se ha venido sujetando mediante "limitadores" especiales, pesadas barras de un metal resistente al calor (de tungsteno o molibdeno, por ejemplo). Sin embargo, los experimentos recientes han demostrado que aun una adición muy pequeña de iones de tungsteno (del orden del 0,1 por ciento) puede dar lugar a una proporción de pérdida de calor por enfriamiento radiativo que disminuye en gran medida la energía de confinamiento en el centro del plasma. El sobrecalentamiento del limitador, tiende pues, a autocorregirse, pues el plasma cambia entonces de ser enfriado por transporte de partículas a ser enfriado por radiación de la impureza. Pero esta clase de remedio no conduce a la obtención de un plasma de reactor con un alto parámetro de confinamiento.

El problema de la interacción del plasma con el limitador puede resolverse

de una manera directa, en el caso de tokamaks con plasmas calentados óhmicamente, en especial a elevadas densidades plasmáticas. El mecanismo de puesta en funcionamiento que se sigue con el tokamak consiste en producir una descarga inicial de gas con una densidad moderada, que luego aumenta de un modo gradual merced a una entrada de gas hidrógeno a pulsos. El enfriamiento superficial del plasma, debido a la afluencia de gas hidrógeno frío en comparación, tiende a proteger el limitador y a suprimir la afluencia de impurezas metálicas. Si el plasma tiene una mezcla importante de iones de impureza ligeros, tales como el oxígeno y el carbono, la radiación de enfriamiento en la superficie del plasma se hace, entonces, más eficaz todavía. De hecho, el enfriamiento superficial en las cámaras de vacío no suficientemente limpias puede llegar a ser con facilidad demasiado eficaz, con la consecuencia previsible de que el canal de descarga del gas se aparte del limitador y continúe contrayéndose hasta que se vuelva muy inestable. En esta última situación, sólo pueden obtenerse bajas densidades de plasma en funcionamiento estable.

El problema de moderar el flujo de calor hacia los limitadores, en tokamaks con fuentes auxiliares de calefacción del plasma, entraña mayor complejidad, sobre todo si se pretende alcanzar en un gran plasma condiciones similares a las de un reactor. Ni tan siquiera una mezcla preparada de iones de impureza ligeros podría aportar una refrigeración radiativa suficiente como para eliminar el flujo de calor fuera del reactor sin enrarecer al propio tiempo el componente de hidrogeniones hasta un nivel incompatible con las exigencias del equilibrio que precisa la energía de fusión. Para la consecución de un reactor quizá haya que acudir a un desviador magnético, un montaje en el que un separador de campo magnético (una línea límite entre dos regiones de diferente topología de líneas de campo magnético) tiene el papel del limitador mecánico ordinario. Otro método para eliminar la gran potencia de calefacción auxiliar entrante sería la creación de un alto nivel de conducción calorífica en una región fría a modo de manta alrededor de un núcleo de plasma caliente y bien confinado.



PROXIMA GENERACION DE TOKAMAKS, representada por el plano del sistema TFTR. Su funcionamiento está proyectado para 1981 y, su instalación, en el laboratorio de Princeton. La cámara de vacío del TFTR tendrá un radio mayor de 265 centímetros y un radio menor de 110 centímetros. El TFTR será el primer aparato magnético de fusión que usará una mezcla combustible de deuterio-tritio, así como el primer sistema capaz de producir energía por fusión en cantidades significativas. El coste estimado para toda la instalación del sistema TFTR se cifra en los 300 millones de dólares.

La inyección de haces intensos de átomos neutros de hidrógeno ha resultado ser el método más eficaz para depositar grandes cantidades de potencia calorífica en el plasma de un tokamak. En este enfoque, desarrollado en los años 60 para aparatos de fusión de extremo abierto, los átomos neutros entran libremente en la botella magnética; luego, a través de choques pierden sus electrones y permanecen en el plasma como iones energéticos. Los iones atrapados se frenan poco a poco y transfieren energía a las partículas del plasma. La aplicación al tokamak de este método de calefacción se realizó en los primeros años de la década de 1970 gracias al desarrollo de fuentes iónicas de gran apertura y alta corriente, capaces de generar haces neutros con energías de decenas de kilovolt en el rango de los 100 kilowatt, en el Laboratorio Nacional de Oak Ridge y en el Laboratorio Lawrence de Berkeley.

En los primeros años de la década de 1970, se aplicaron haces neutros en el tokamak ATC de Princeton y en el Ormak de Oak Ridge. Se detectaron incrementos de varios cientos de electronvolt en la temperatura iónica, y en aquel entonces se consideró que marcaban el comienzo de una nueva era de la calefacción no-óhmica del tokamak. Por desgracia, durante ese proceso de calentamiento, los electrones del plasma solían enfriarse, debido, al parecer, a la inyección de los iones de impureza motivada por el impacto del haz de partículas sobre el limitador y sobre las paredes de la cámara de vacío. En 1977, se inyectaron potencias de medio megawatt; las temperaturas iónicas se habían elevado a dos keV en el Ormak y esa misma cifra en el aparato TFR de Fontenay-aux-Roses (Francia). Pero los electrones seguían mostrándose reacios. Empeza-

ron a observarse pequeñas elevaciones de la temperatura electrónica, aunque con la suficiente claridad como para hacer posible la fijación de la dependencia de la temperatura con la energía de confinamiento.

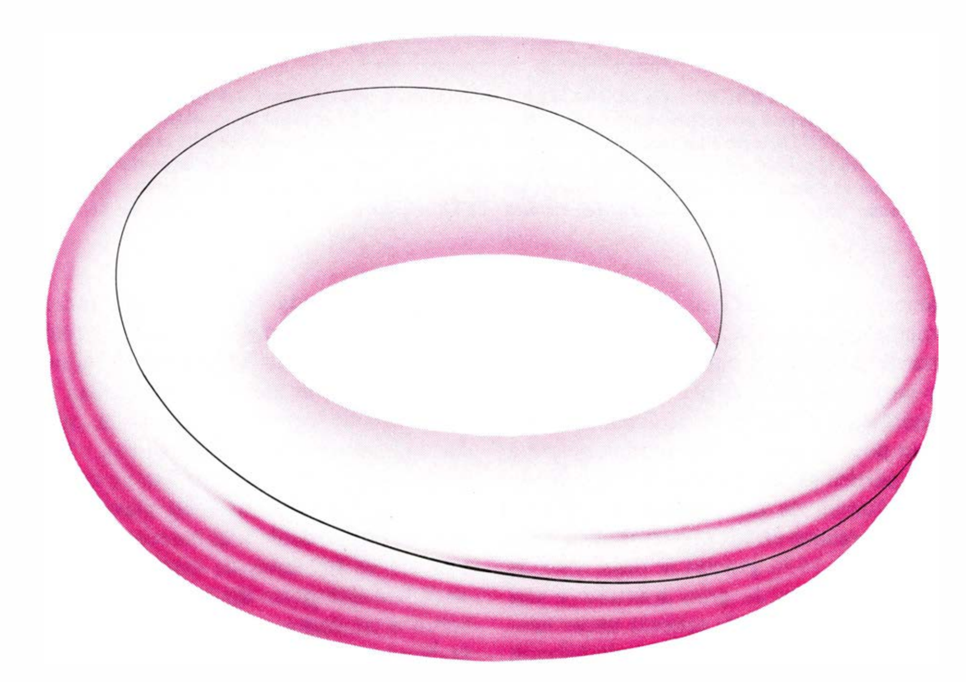
Entretanto, habían hecho su aparición dos nuevos tokamaks, ambos con un plasma de radio menor de unos 40 centímetros: el T-10, construido bajo la dirección de V. S. Strelkov en el Instituto Kurchatov, y el PLT, construido bajo la dirección de Wolfgang Stodiek, en Princeton. En la operación de calentamiento óhmico, estos gigantescos tokamaks tuvieron el esperado debilitamiento de la densidad de potencia asociado a su tamaño, por lo que hubo que limitar su temperatura iónica máxima del kiloelectronvolt, aproximadamente, y la densidad del plasma a algo más de 10^{14} partículas por centímetro cúbico. Se vio que los tiempos de confinamiento de energía aumentaban con el cuadrado del radio menor, que era lo esperado, y que alcanzaban valores de alrededor de 100 milisegundos, pero los parámetros de confinamiento asociados apenas si habían ascendido por encima de 10^{13}. La verdadera ventaja de los grandes tokamaks ha resultado ser su accesibilidad para entradas de calor externo de alta potencia.

En el otoño de 1977, un grupo dirigido por Harold P. Eubank puso en funcionamiento, en el PLT, dos líneas de haces neutros que trabajaban con una energía de inyección de 40 keV. Estos inyectores, que formaban la primera entrega de un lote de cuatro, se desarrollaron y construyeron en las instalaciones del PLT en Oak Ridge; estaban capacitados para inyectar una potencia total de 1,1 megawatt de haz neutro en el plasma. En diciembre, la temperatura iónica del PLT subió hasta 2,2 keV –un nuevo récord por muy escaso margen–, aunque los iones quedaban retenidos mediante choques con electrones más fríos, resultando de ello pérdidas radiativas para los iones de impurezas metálicas. En estas condiciones se lograron varios descubrimientos importantes en la espectroscopía del tungsteno y del hierro altamente ionizados, pero la oportunidad de un avance en la temperatura iónica parecía limitada, incluso en presencia de energías de inyección más altas.

Utilizando grafito como limitador se hicieron algunos experimentos exploratorios que abrieron grandes esperanzas. El carbono chisporrotea más fácilmente que el tungsteno, pero una mezcla de iones de carbono es mucho menos perjudicial para el equilibrio energético del plasma debido a que el carbono se ioniza completamente en el interior del plasma e irradía principalmente desde su borde frío. A principios del pasado año, las dos líneas de haces neutros que faltaban fueron ajustadas al PLT, al mismo tiempo que se incorporaba un nuevo conjunto de limitadores de grafito refrigerados por agua. En julio se inyectó en el PLT un haz neutro de dos megawatt de potencia y finalmente la temperatura electrónica respondió elevándose a 3,5 keV. Bajo estas condiciones, la temperatura iónica pudo subir a 5,5 keV. Este pico de temperatura iónica equivalente a más de 60 millones de grados C, mejoraba sustancialmente los cálculos previstos y tuvo un atractivo especial para los investigadores de la fusión: por primera vez, en un aparato tokamak se había superado la temperatura iónica mínima necesaria para la ignición en un plasma de deuterio-tritio.

Sin embargo, el principal interés científico se centraba en el incremento del confinamiento del plasma. En los experimentos de calefacción del PLT se utilizaba una densidad en el centro del plasma comparativamente baja, de unas 5×10^{13} partículas por centímetro cúbico; en efecto, se puede conseguir una temperatura de las partículas más alta reduciendo el número de partículas que se reparten la potencia calorífica entrante. En este régimen de alta temperatura y baja densidad, el plasma del PLT logró, por primera vez, un grado de no colisión similar al de un reactor tokamak, sin que se encontrase ningún deterioro observable en el confinamiento. La magnitud de la conducción de calor por los iones continuó siendo aproximadamente consistente con las predicciones de la teoría de la difusión de Coulomb, en lugar de mostrar el señalado crecimiento predicho por la teoría de la partícula atrapada en microinestabilidades. Como premio de consolación para los teóricos de la microinestabilidad, el plasma del PLT empezó a mostrar un nuevo tipo de fluctuación, de alta frecuencia, de la densidad a temperaturas más altas. Diagnósticos detallados de la fluctuación podían revelar todavía que era realmente del tipo predicho. Sin embargo, el nuevo fenómeno no tenía ningún impacto medible en el confinamiento real del plasma; esta observación mantiene su verdad aún cuando la temperatura iónica ha sido elevada entretanto hasta 6,5 keV (75 millones de grados C.) con una potencia del haz neutro inyectado de 2,5 megawatt.

El notable aumento de la temperatura electrónica producido por calefacción no óhmica de haz neutro en el PLT ha posibilitado una primera aproximación a la dependencia real entre la temperatura y la conducción de calor por parte de los electrones. A primera vista, no había separación entre el tiempo de confinamiento de energía para los electrones y el valor prescrito por la simple ley empírica de dependencia observada en prue-



PROXIMA GENERACION DE PROBLEMAS que se prevén en el terreno de la investigación sobre el tokamak. Se espera topar con una nueva clase de inestabilidades macroscópicas, que se les ha dado el nombre de "modos de intercambio de presión", de los que se avanza uno aquí. En los experimentos acometidos, una elevada razón entre la presión cinética del plasma y su presión de campo magnético, el plasma puede presentar una tendencia a comprimir las líneas de campo de la botella magnética deformándolas ligeramente; resultaría un efecto de rizados alrededor de la superficie exterior del toro plasmático, alineado, con cierto parecido, con el paso de rosca de las líneas de campo magnético.

bas efectuadas en el Alcator *A* y otros tokamaks calentados óhmicamente: el tiempo de confinamiento resultaba proporcional a la densidad multiplicada por el cuadrado del radio menor del plasma. Pero una inspección más ajustada de la evolución del perfil de la temperatura electrónica durante la calefacción por haz neutro nos ha revelado el resultado sorprendente y alentador de que la conducción calorífica por los electrones decrece en la parte más caliente del núcleo del plasma, mientras que permanece aproximadamente igual en los bordes. Está, pues, al caer el establecimiento de una ley empírica de dependencia explícita, ahora que podemos variar la temperatura electrónica de un tokamak, independientemente del campo magnético poloidal y de la densidad de partículas del plasma. Cuando se llegue a esa ley, podremos optar entre los modelos teóricos de conducción de calor.

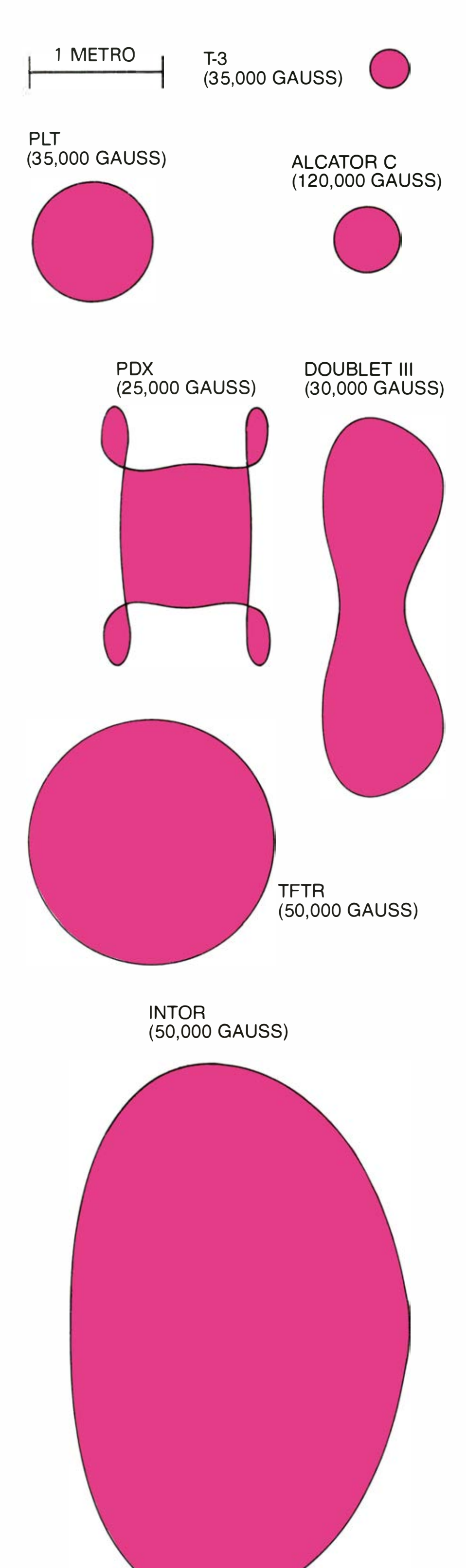
¿Está muy lejos, el régimen del modelo tokamak, de cumplir los requisitos de un reactos capaz de producir potencia a gran escala? Puesto que el dominio de temperatura de tal reactor se introdujo en el PLT a baja densidad de funcionamiento, el parámetro de confinamiento fue sólo de unos 10^{12}. Para elevar la densidad y el tiempo de confinamiento del plasma se necesitarán mayores potencias de haz neutro y energías de los iones inyectados más elevadas. Sin embargo, aun dentro del presente régimen del PLT, la liberación de potencia de fusión ya es importante. Cuando se utilizan plasmas y haces de deuterio, se producen alrededor de 4×10^{13} fenómenos de fusión en cada pulso. La reacción de deuterio-tritio no es utilizada en la serie de experimentos corrientes en la física de plasmas, porque los problemas de seguridad radiológica asociados con el manejo del tritio complicarían innecesariamente los procedimientos experimentales. En su substitución, se calcula el equivalente de deuterio-tritio sobre la base de las razones conocidas en las reacciones de fusión, que son unas 300 veces más grandes para una mezcla combustible de deuterio-tritio que para el deuterio puro. En el PLT la potencia de fusión de deuterio-tritio equivalente es de unos 50 kilowatt, correspondiendo a un valor de *Q* de 0,02, hasta la fecha un verdadero récord.

El valor de *Q*, comparativamente alto, obtenido en el PLT se atribuye a la naturaleza del método de calefacción. La calefacción por haz neutro introduce la potencia de entrada del modo más favorable: sube la "cola" de elevada energía en la distribución energética de los iones, responsable de la mayor parte de la reactividad por fusión. Debido a este hecho especial, la energía iónica media en el centro del plasma del PLT con una temperatura iónica de 6,5 keV es de unos 13 keV (contando los iones de deuterio inyectados) y la fusión que se tiene es unas 100 veces mayor que la que se consigue cuando se usa para calentar un haz de hidrógeno no reactivo.

Desde la perspectiva de los años 50, cuando eran normales temperaturas del plasma inferiores a los 100 electronvolt y parámetros de confinamiento del orden de 10^{10}, no resultaban de especial importancia ciertos detalles del proyecto de reactores tales como el grado de separación entre una curva de distribución de energía y la distribución normal de Gauss. Se pretendía superar la combinación mágica de una temperatura de 10 keV con un parámetro de confinamiento de 6×10^{13}, umbral comúnmente conocido como el criterio de Lawson para plasmas de deuterio-tritio (en honor del físico J. D. Lawson). Pero hoy, cuando los experimentos tienen por meta iniciar el régimen del reactor, la perspectiva ha cambiado. Los pormenores del reactor toroidal importan, y mucho, y se les afronta ya en función del material y del coste.

La estructura fundamental de todos los reactores de deuterio-tritio viene a ser la misma. Alrededor del 80 por ciento de la potencia de fusión saldría del plasma en forma de neutrones de 14 Mev y serían capturados como calor en una capa de litio, que serviría para "regenerar" el suministro de tritio al plasma. El calor generado por los neutrones, unido al flujo de calor saliente del propio plasma, se convertiría luego en energía eléctrica en un ciclo de vapor convencional. Otra característica común proyectada es que las bobinas magnéticas del reactor tendrían que ser superconductoras, puesto que el consumo de potencia por las bobinas de cobre ordinarias habrá de superar, de acuerdo con los cálculos, la capacidad del reactor de generar energía eléctrica.

Entre los diferentes proyectos para los reactores de deuterio-tritio, las notas distintivas de cada uno tienen mucho que ver con las características del propio plasma. En los reactores toroidales hay una jerarquía de modos de funcionamiento posibles, estimados por orden de los valores crecientes de *Q*. En los regímenes donde pueden obtenerse con más facilidad las partículas, éstas tienen distribuciones especiales de la energía que son comparativamente ricas en iones energéticos. En este caso deberían conseguirse valores de *Q* del orden de 2 o 3,



SECCIONES transversales menores de distintos plasmas tokamak, desde el T-3 hasta el INTOR, dibujadas todas a la misma escala. En cada aparato se indica la intensidad de campo magnético toroidal. El PDX (mostrado aquí en su forma casi cuadrada) y el tokamak Doublet III, de la General Atomic Company, están proyectados para utilizar la inyección de haz neutro de alta energía al objeto de investigar las inestabilidades inducidas en la presión de los tokamaks con distintas configuraciones de la sección transversal menor. La forma de *D* alargada es la privilegiada, hoy, para la sección transversal menor del INTOR.

con parámetros de confinamiento muy bajos, 10^{13}. Por desgracia, en un reactor con un valor de Q tan pequeño las desventajas son de tipo económico: una parte substancial de la energía eléctrica de salida debería emplearse para mantener caliente al plasma y para producir energía neta se necesitarían sistemas de alto rendimiento. Por último, la solución atractiva sería la de un reactor que funcionase con un régimen de ignición o de casi ignición, donde la mayor parte de la calefacción del plasma correría a cargo de las partículas alfa producidas en la reacción deuterio-tritio. En esta hipótesis las partículas del plasma tenderían a mostrar una buena distribución de la energía por colisiones de Coulomb y el parámetro de confinamiento tendría que llegar a unos 3×10^{14}. Ya que se espera que aumente el tiempo de confinamiento de energía con el cuadrado del radio menor del plasma, ello significa que el tamaño mínimo de un plasma de tokamak en ignición tiende a ser varias veces mayor que el de un plasma de tokamak con un valor de Q pequeño.

La obtención de valores de Q algo mayores que la unidad es una perspectiva realista para la próxima generación de experimentos con el tokamak. Dos de los grandes tokamak que actualmente están en fase de construcción lograrán demostrar este tipo de funcionamiento basado en pulsos cortos en plasmas reales de deuterio-tritio. Estos aparatos son el Reactor de Prueba de Fusión Tokamak (TFTR del inglés "Tokamak Fusion Test Reactor") en Princeton, cuya terminación está prevista para finales de 1981, y el Toro de la Unión Europea (JET "Joint European Torus") en Culham, Inglaterra, terminado en 1983; ambos tendrán corrientes de plasma de unos tres millones de ampère, potencias de calefacción por haz neutro de decenas de megawatt; y se espera que generen por fusión energías de salida entre 10 y 100 megawatt. El radio menor del plasma en el TFTR y en el JET serán, respectivamente, dos y tres veces superiores al radio menor del plasma del PLT. Se estima que el coste de cada una de las nuevas instalaciones será del orden de los 300 millones de dólares.

En principio se podría intentar progresar "de una manera directa" a partir de esta próxima generación de experimentos hasta un ciclo de servicio de alta potencia que produjera la ignición de un reactor en el rango de los 100 a los 1000 megawatt, mediante el aumento del tamaño del plasma y la adición de una capa de litio y de bobinas magnéticas superconductoras. Sin embargo, el coste que se estima para este paso es de unos mil millones de dólares. Además, quedan por resolver muchos problemas científicos y técnicos importantes. Para afrontar estos problemas con la máxima eficacia, los principales contribuyentes mundiales en la investigación sobre fusión —los Estados Unidos, la Unión Soviética, Europa Occidental y el Japón— han constituido un grupo de estudio llamado INTOR (del inglés "International Tokamak Reactor") bajo los auspicios de la Agencia Internacional de Energía Atómica en Viena. Dentro del programa de los Estados Unidos sobre fusión existe un grupo análogo llamado Instalación de Ingeniería de Prueba (ETF, del inglés "Engineering Test Facility"), con sede en Oak Ridge. Se confía que estos estudios presenten anteproyectos avanzados y ponderados de un reactor tokamak en ignición, que puedan llevarse a la práctica a principios de la década de 1990, así como un esquema del programa de las tareas de investigación y desarrollo que faltan por ser tratadas.

Un punto clave, dentro de la última consideración, es maximizar la presión cinética del plasma del tokamak con respecto a la presión del campo magnético. Se espera que la razón de presión que se consiga (parámetro llamado valor beta del plasma) esté limitada por una clase de inestabilidades magnetohidrodinámicas macroscópicas conocidas por modos de intercambio de presión ("ballooning interchange modes"), en las cuales el plasma se comprime entre las líneas de campo de la botella magnética al tiempo que las deforma ligeramente. Cuanto menor sea el valor beta, menos podrá el plasma deformar el campo magnético y más confianza merecerá el confinamiento del plasma. Por contra, cuanto mayor sea el valor beta, tanto más elevada será la densidad del plasma a una temperatura determinada, mayor será asimismo el parámetro de confinamiento y mayor también la producción prevista de potencia por fusión en un plasma de un tamaño dado.

Hoy estamos asistiendo a un fuerte empeño teórico para proyectar la geometría óptima del tokamak y estimar el máximo valor beta que pueda conseguirse, usando alguno de los mayores ordenadores del mundo. Los cálculos actuales del valor beta van desde el 3 al 10 por ciento, que, en general, sería adecuado para un reactor, pero depende mucho de si debe aplicarse el extremo superior o el inferior del intervalo. Para alcanzar valores beta superiores al dos por ciento en los experimentos actuales, hay que emplear calefacción auxiliar suficientemente intensa; de acuerdo con lo anterior no se ha encontrado todavía ninguna inestabilidad de fuga de presión. Sobre este punto es probable que nos lleguen, en los dos próximos años, las primeras noticias significativas del tokamak ISX de Oak Ridge, del PDX de Princeton y del Doublet III del laboratorio de la Compañía General Atómica en San Diego, todos los cuales utilizarán inyección de haz neutro de alta potencia para determinar los valores beta de tokamaks con formas distintas de la sección transversal menor. La candidata hoy es una sección transversal alargada en forma de *D*, pero cuál ganará y cuál será la magnitud de su valor beta son incógnitas todavía.

Una segunda cuestión de fundamental interés en el proyecto de un reactor tiene que ver con el mejor método para controlar las impurezas. La aplicación del desviador magnético, que está a punto de ser comprobado en el aparato PDX, parece ser una buena opción, pero agrandaría enormemente el tamaño total del reactor. La elección definitiva se aclarará en 1982, cuando empiece a funcionar el JT-60, un proyecto japonés de tokamak en la línea de los aparatos TFTR y JET; el JT-60 no quemará una mezcla combustible de deuterio-tritio, pero, a diferencia de los dos gigantes, tendrá un sistema de desviación magnética.

Se espera que el problema de mantener una composición del plasma deseable llegue a ser de particular gravedad cuando los experimentos del tokamak pasen del actual ciclo de servicio característico de aproximadamente un pulso de un segundo de duración cada 10 minutos a pulsos que duren 10 minutos o más, separados por intervalos mucho más breves. Será necesaria una nueva generación de tokamaks con bobinas magnéticas superconductoras para afrontar el régimen de alto rendimiento. El primer paso importante en esta dirección lo están dando los rusos con su aparato T-15, que entrará en funcionamiento en el Instituto Kurchatov en 1983. El desarrollo de métodos de calefacción del plasma más sencillos que el de la inyección de haz neutro continúa siendo otro medio de progreso técnico que condicionará el destino de los reactores tokamak. El calentamiento del plasma por microondas, del que se espera mucho tecnológicamente, es objeto de examen en un amplio frente en los actuales experimentos con tokamak.

El éxito, constante y creciente, del tokamak para abordar el tema de la fusión habría complacido a los comités del Congreso norteamericano que revisaron

el estado de la investigación sobre fusión en los años 60. Trataron de acotar el centro de la investigación para desechar las alternativas menos prometedoras. Ahora que se ha conseguido un centro más perfilado, surge una nueva preocupación: ¿Debería privilegiarse una solución particular sobre otras aunque éstas hayan pasado la prueba crítica? El reactor tokamak tiene garantías de funcionamiento, pero su tamaño descomunal y su compleja geometría pueden, en último término, limitar su atractivo económico frente a las máquinas espejo, reactores inerciales y toda una gama de otros enfoques que parecen reclamarse en función del potencial económico, de interés creciente, y de las decrecientes pruebas técnicas.

Aunque este tipo de argumentación resulte quizá paradójico, tiene un punto válido. La abundancia de alternativas tecnológicas en el campo de la investigación sobre fusión representa un recurso valioso, que debería cultivarse con más atención que nunca cuando los primeros reactores experimentales de fusión sean una realidad. Aunque el tokamak constituye hoy la mejor solución para el primer reactor, quizá la industria de la fusión acabe por adoptar otros modelos que nos parecen ahora puras quimeras. A largo plazo, las posibilidades de evolución son vastísimas.

La aparición a principios del siglo XXI de una economía energética basada en los reactores de fusión tokamak que quemen una mezcla combustible de deuterio-tritio permitiría que la energía nuclear gozara de mejor prensa ecologista y ganara en aceptación por parte del gran público. En un reactor de deuterio-tritio no cabría una reacción incontrolada, tendría poco calor procedente de la radiactividad y produciría residuos radiactivos con un riesgo de desastre y una vida muy menguados. Sin embargo, parece probable que, en la primera o dos primeras centurias de energía por fusión comercial, la reacción de deuterio-tritio podría sustituirse por procesos de fusión más difíciles de conseguir, pero que ofrecerían menos riesgos ambientales: primero la reacción deuterio-deuterio y luego una de las reacciones de fusión más exóticas, en particular la que libere todo su calor en forma de partículas cargadas, en lugar de neutrones. La adaptación de la energía de fusión a las necesidades sociales y a las exigencias del medio ambiente no está coartada por ninguna restricción física, sino por un conocimiento científico imperfecto y por los límites de nuestra imaginación tecnológica.

Una aldea de colonos preneolíticos a orillas del Eúfrates

El período neolítico está asociado tradicionalmente con el nacimiento de la agricultura. Pero antes, en Siria, algunos de los primeros aldeanos hicieron notables experiencias de granja, al tiempo que cazaban y recolectaban

Andrew M. T. Moore

El estadio de la prehistoria humana en que el hombre adoptó por vez primera una economía agrícola es de un interés fundamental para los arqueólogos, porque representa la divisoria entre el modo de vida cazador-recolector de nuestros antepasados del período paleolítico, o Edad de la Piedra Antigua, y la primera forma de vida civilizada. Durante generaciones, los estudiosos han explorado el suroeste de Asia en un esfuerzo por averiguar exactamente dónde y cuándo dio comienzo la agricultura. El ritmo de la exploración y de la excavación se ha acelerado notablemente en los treinta años últimos, pero ahora a los arqueólogos les preocupa menos dónde y cuándo se desarrolló en esta zona el modo de vida agrícola que el cómo y el por qué del mismo. Un reciente trabajo de análisis en un yacimiento del valle del Eúfrates en el norte de Siria ha proporcionado elementos para responder con suficiencia de datos a estas dos últimas preguntas.

La zona del suroeste asiático que encierra mayor número de estaciones conocidas y excavadas de las épocas epipaleolítica (la que sigue inmediatamente al Paleolítico) y neolítica es el área conocida tradicionalmente como levante asiático. Hoy en día toda esta área se halla repartida entre las naciones modernas de Siria, el Líbano, Jordania e Israel. La estación donde se ha realizado el trabajo que voy a describir se halla en el norte de Siria. Se trata del mayor yacimiento del Neolítico inicial, con mucha diferencia sobre todos los demás del Próximo Oriente, y es un *tell*, una extensa loma formada por los restos de varios milenios de ocupación humana. La loma en cuestión, conocida como Tell Abu Hureyra, estaba destinada a quedar sumergida cuando el agua del Eúfrates se viese retenida por una nueva presa que se estaba construyendo a unos kilómetros aguas abajo, en Tabqa. Vi la loma por primera vez en marzo de 1971, tres años antes de que estuviese destinada a desaparecer, y me quedé impresionado por su tamaño y por su interés potencial para la investigación.

Era evidente que el grueso depósito de detritus, que medía hasta ocho metros de altura en una extensión de 11,5 hectáreas, debía haberse formado durante un largo periodo. Cabía por tanto la esperanza de que unas excavaciones en Abu Hureyra proporcionasen abundantes testimonios acerca de los comienzos de la vida sedentaria en esta frontera oriental del Levante, donde hoy sólo caen unos 200 milímetros anuales de lluvia en lo que es una árida estepa.

La superficie de la loma estaba cubierta de una densa lluvia de utensilios de pedernal y de desperdicios de un tipo semejante al encontrado años antes en las excavaciones del yacimiento neolítico de Buqras, aguas abajo del Eúfrates, por Henri de Contenson, del Centre National de la Recherche Scientifique francés (C.N.R.S.). Buqras fue datado alrededor del 6000 a.C. (Todas las fechas que voy a dar aquí se fundan en cálculos del carbono-14 sacados de la vida media de ese isótopo radiactivo, pero sin someterlos a la calibración.) Esto sitúa Buqras hacia mediados de la época neolítica del Próximo Oriente, donde la Edad de la Piedra reciente empezó hacia 8500 a.C. y continuó en la mayor parte de la región hasta fechas tan bajas como entre 4000 y 3500 a.C. Si en la superficie de Abu Hureyra había utensilios como los de Buqras, era de esperar que debajo de la superficie se encontrasen restos del Neolítico arcaico, o más primitivo, y acaso por debajo de éstos, restos del Mesolítico que lo precedió, el cual dio comienzo aquí hace unos 20.000 años y continuó hasta 8500 a.C., aproximadamente.

Mis colegas y yo empezamos a trabajar en Abu Hureyra en 1972, bajo el patrocinio del Museo Pitt Rivers de la Universidad de Oxford y con la ayuda de una serie de instituciones, tanto del Reino Unido como de Norteamérica. Al término de la primera campaña de excavación, nuestros más optimistas afanes se habían visto satisfechos. Bajo los niveles de ocupación del Neolítico arcaico y pleno, que databan desde alrededor del 7500 a.C. hasta una fecha algo posterior al 6000 a.C., aparecían señales de una ocupación mesolítica del yacimiento, que se había prolongado durante varios siglos a finales del X y comienzos del IX milenio a.C. Otra campaña, en 1973, puso fin a la excavación y la gran loma fue sumergida a principios de 1974, tal y como estaba programado, dejando a nuestro grupo dueño de un inmenso montón de artefactos y de restos de animales y plantas para ulterior examen. Sólo ahora, cuando está en marcha el estudio del material, nos percatamos de la verdadera importancia de la excavación por lo que se refiere a la transición de un modo de vida cazador-recolector a otro agrícola.

La zona de ocupación mesolítica de Abu Hureyra se halla en el extremo norte de la loma, en el borde de la llanura aluvial del Eúfrates. Consiste en un simple poblado aldeano: una serie de fondos de cabaña y otros huecos excavados en la superficie del suelo original que constituye la base del *tell*. Cada fondo de cabaña tenía una armadura de pies derechos de madera; los techos, de los que no queda ninguna señal, estaban probablemente construidos con juncos. Los fondos y huecos se fueron llenando gradualmente de restos de ocupación, que se acumularon en un espesor de un metro, más o menos. Junto a los restos había muchos de los artefactos que los ocupantes mesolíticos habían empleado.

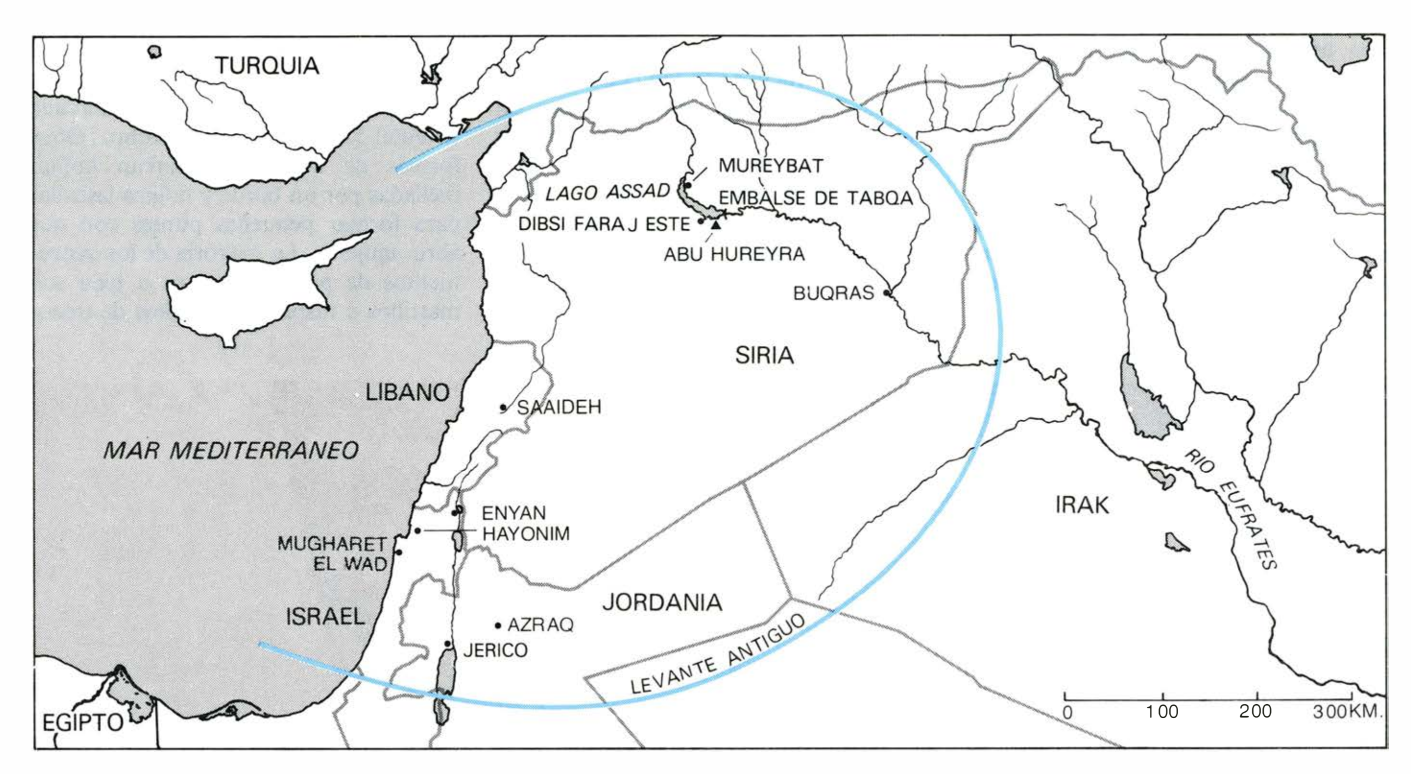
Entre los artefactos llamaban la atención morteros de piedra, raspadores y piedras de moler. Muchas de ellas estaban manchadas de ocre rojo, prueba de que habían sido usadas para triturar y convertir en polvo este pigmento decorativo. La presencia de los instrumentos de moler en los fondos de cabaña y en los huecos, indica que también fueron empleados para preparar alimentos de origen vegetal.

Los artefactos hallados en mayor número eran instrumentos de sílex de dos tipos. Muchos son tan pequeños que se les llama microlitos, pero hay un cierto número de objetos de sílex más pesados. También encontramos enormes cantidades de desechos de sílex.

Los microlitos más corrientes, con mucha diferencia, son unas cuchillas pequeñas en forma de crecientes y conocidas como lunuladas; estas hojitas pudieron ir encajadas en astiles de madera para servir de aletas de una flecha, de filo de un cuchillo o (como se da el caso en otros puntos del Cercano Oriente) como dientes de una hoz. Otras formas de microlitos abarcan hojitas melladas por un borde y hojitas lascadas para formar pequeñas puntas con que abrir agujeros. La mayoría de los instrumentos de piedra mayores o bien son martillos o raspadores, hechos de trozos



DOS NIVELES DE OCUPACION de Tell Abu Hureyra aparecen en esta fotografía de una calicata realizada en el extremo norte de la loma de 11,5 hectáreas. Las paredes, de adobes rectangulares, de los aldeanos del Neolítico arcaico que se asentaron en el yacimiento del río Eúfrates hacia 7500 a.C. son visibles en la parte superior. Sepultada baja esta aldea neolítica, había una aglomeración de fondos de cabañas y huecos de trabajo usados por cazadores-recolectores del período mesolítico tardío hasta aproximadamente 8500 a.C. Estos cultivaban evidentemente grano. Los agujeros para postes que rodean los fondos sostenían los soportes de paredes y techos, fabricados con juncos. Los agujeros pequeños fueron hechos por roedores.

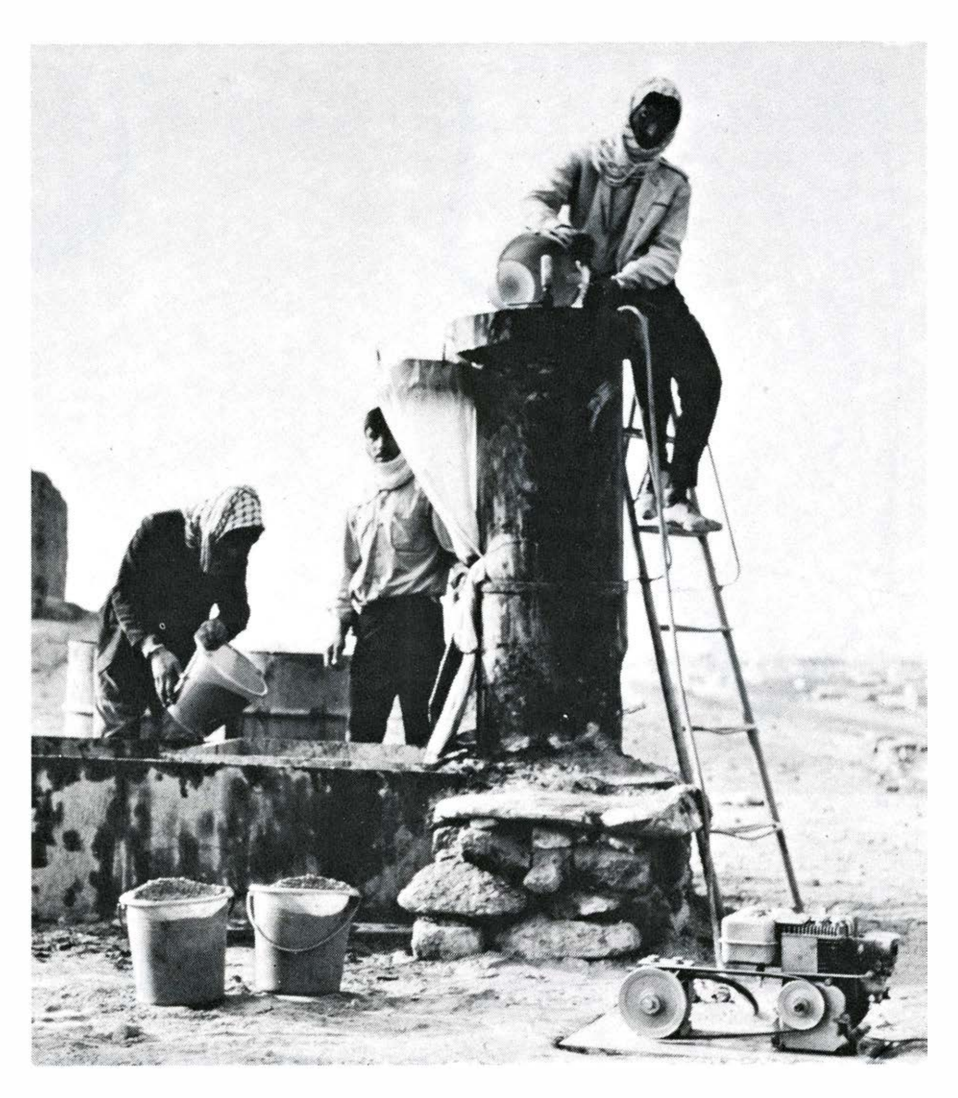


LEVANTE ASIATICO en la Antigüedad, una zona repartida ahora entre los estados modernos de Israel, el Líbano, Siria y Jordania. Constituye la parte del suroeste de Asia que encierra el mayor número de yacimientos mesolíticos y neolíticos excavados. Abu Hureyra quedó sumergida en 1974.

LA LOMA DE LA ALDEA se halla en una terraza baja, al sur de la llanura aluvial del Eúfrates. El Wadi Hibna pudo haber sido un arroyo permanente en el Neolítico arcaico, que suministrase agua para regadíos. Los números de las curvas de nivel corresponden a metros sobre el mar.

de pedernal de cierto tamaño. Los artesanos de esta fase mesolítica también fabricaron laboriosamente platos de piedras distintas del pedernal, e hicieron muescas en los lados de guijarros, posiblemente para adaptarlos al uso de pesas de redes de pescar.

También se encontraron en abundancia utensilios hechos de hueso. La mayoría de ellos son perforadores para trabajos duros; otros artefactos de hueso comprenden espátulas, agujas y puntas de dos cabos que pueden haber sido puntas de flechas. Entre los artículos de carácter decorativo había cuentas y colgantes de hueso, junto con cuentas y colgantes hechos de concha y de piedra.



MAQUINA DE FLOTACION diseñada en la Universidad de Cambridge. Fue una de las dos máquinas usadas por los excavadores de Abu Hureyra para extraer unos 1000 litros de restos de plantas, de varios niveles del yacimiento. Los niveles mesolíticos encerraban tres cereales: trigo, centeno y cebada.

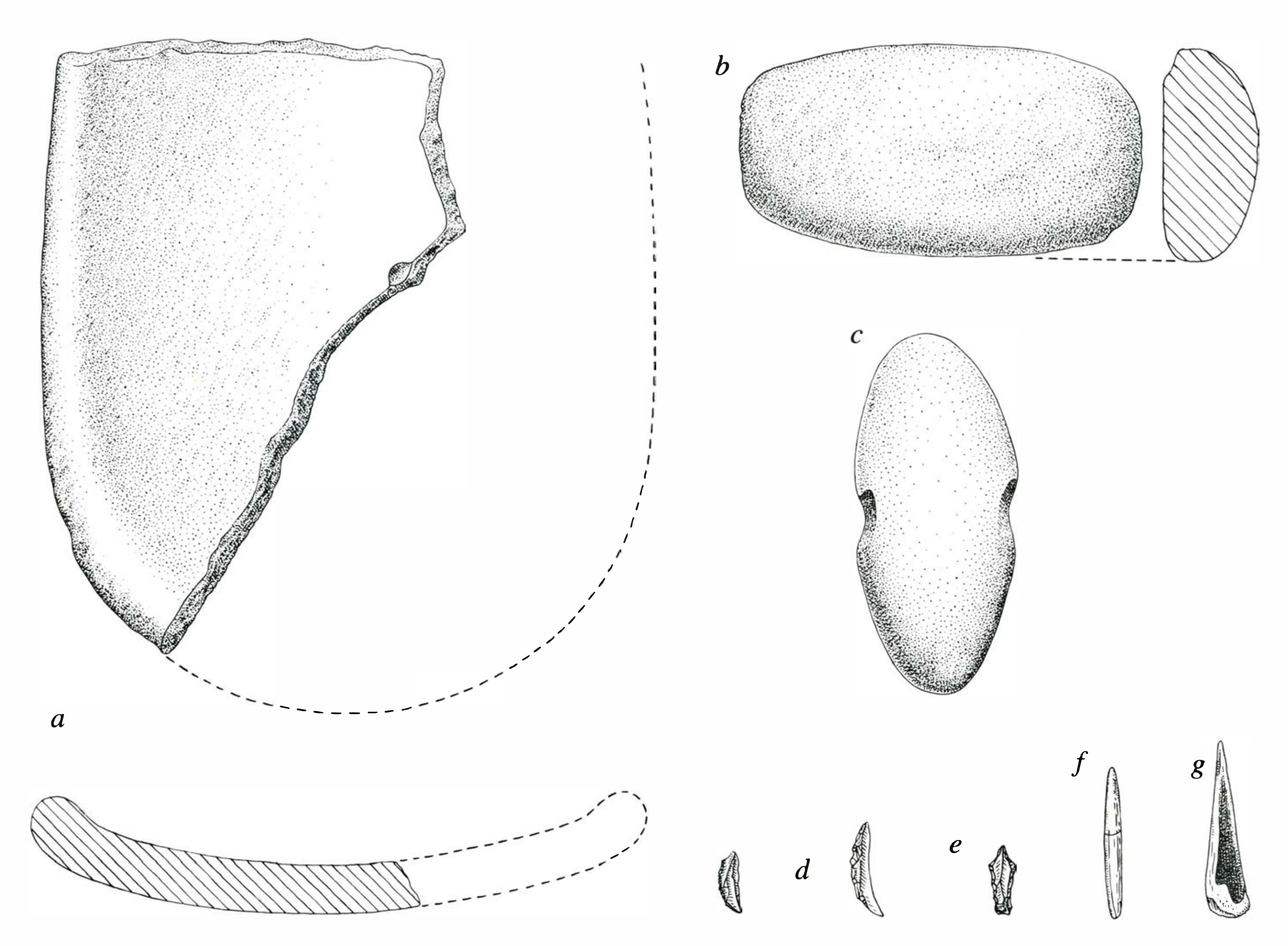
Los artefactos asociados con la aldea mesolítica de Abu Hureyra son típicos de la fase tardía de este periodo en otros lugares del Cercano Oriente, aproximadamente contemporáneos de yacimientos, tales como los de la cultura natufiense de Palestina en Mugharet el Wad, Eynan y Hayonim. Durante el último periodo mesolítico la población humana del Cercano Oriente era considerablemente mayor de lo que había sido durante los varios milenios precedentes. El incremento de población parece haber correspondido a una notable mejoría del ambiente. Mientras la fase anterior se había caracterizado por condiciones frescas y secas, el estadio posterior, y el periodo neolítico que siguió, presenciaron un ascenso gradual de la temperatura y un aumento de las precipitaciones.

La consecuencia para esta zona fue una expansión hacia oriente de la zona de bosque mediterráneo que anteriormente había estado limitado a las montañas costeras de la región. Esta área boscosa acabó por alcanzar el borde de la meseta interior. La estepa situada al este y al sur del área forestal aumentada quedó separada del bosque propiamente dicho por una zona intermedia de arbolado abierto. Durante el periodo de tiempo cada vez más cálido y húmedo, la estepa misma se hizo más rica en especies vegetales y animales. Como consecuencia, la zona intermedia entre el bosque denso y la estepa llegó a brindar oportunidades particularmente atractivas, que las poblaciones humanas en vías de practicar la agricultura eran libres de explotar.

Abu Hureyra, situada en el borde de la estepa, parece haber sido una de tantas estaciones que nacieron como consecuencia de la mejoría del medio ambiente. El proceso es fácil de reconstruir. La mayor pluviosidad traería consigo una mayor abundancia de vegetación, que para pequeños grupos de cazadores-recolectores mesolíticos habría sido un factor de importancia capital. Hoy diríamos que la vegetación, más abundante, pudo sostener una biomasa mayor; y desde luego esto se halla atestiguado por el incremento, demostrable, de la población humana. Evidentemente, algunos de los grupos de cazadores-recolectores comenzaron a reunirse y a implantar un género de vida sedentario, o por lo menos semisedentario. La concentración de tales grupos acrecentados, en estaciones tales como Abu Hureyra, creó por sí misma nuevas circunstancias; por ejemplo, relaciones sociales más ricas. En tales circunstancias era ventajoso desenvolver nuevas estrategias de subsistencia en apoyo de la nueva existencia sedentaria.

Tal parecía, por lo menos, una hipótesis plausible para explicar la aldea mesolítica que descubrimos en Abu Hureyra. Para poner a prueba la hipótesis, los excavadores se esforzaron en recoger todos los posibles testimonios de la economía del poblado. Esto requería técnicas de excavación rara vez empleadas antes en Oriente. En primer lugar, trabajando con cedazos de malla fina, los excavadores tamizaron toda la tierra que habían levantado de la excavación. Mediante la técnica del tamizado se recobró una gran cantidad de huesos de animales, más de dos toneladas en conjunto. En segundo lugar, se aplicó un método de flotación para separar las finas materias vegetales de la tierra tamizada. Dos máquinas de flotación, diseñadas en la Universidad de Cambridge, recuperaron sistemáticamente unos 1000 litros de restos de plantas de los varios estratos de la excavación.

El examen de estos restos de plantas, efectuado por Gordon Hillman, de la Universidad de Gales, revela que los habitantes mesolíticos de Abu Hureyra recolectaban, si de hecho no las cultivaban, varias leguminosas, incluido un tipo silvestre de lenteja. Recogían pequeñas nueces del árbol del terebinto (un



ARTEFACTOS MESOLITICOS: piedras de moler, de basalto (*a*); trituradoras (*b*) y guijarros con muescas, tal vez pesas de red (*c*). Entre los microlitos de retoque tallado, los lunulados (*d*) eran los más abundantes; los seguían los microburiles (*e*). Los útiles de hueso incluyen puntas de dos cabos (*f*), posiblemente para flechas, y muchos buriles (*g*). Las piedras de moler eran usadas para pulverizar ocre y quizá para triturar plantas alimenticias.

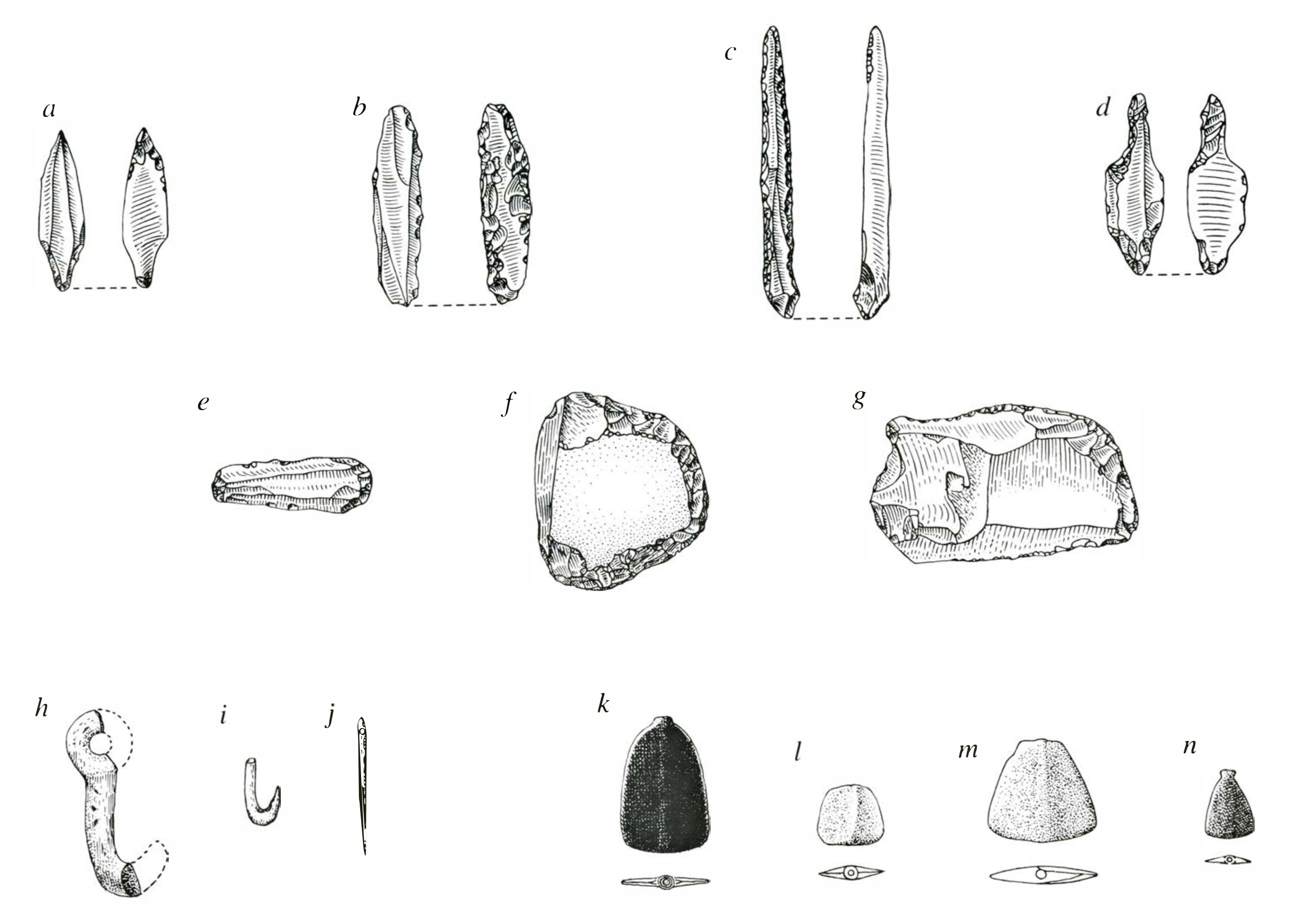
pariente del pistacho), frutos de arbustos de la familia del olmo, alcaparras y semillas del espolín silvestre. Lo que constituyó una apasionante y significativa sorpresa fue el que Hillman descubriese granos de cereal. La mayoría de ellos son granos de un trigo primitivo: la escanda silvestre. Otros dos cereales están presentes en menor abundancia: las formas silvestres del centeno y de la cebada. Como el cultivo de los cereales es uno de los jalones iniciales de la revolución neolítica, el descubrimiento de las tres clases de cereal en los niveles mesolíticos de Abu Hureyra planteaba un problema crucial: ¿Alguno de los tres (y quizá también las leguminosas) eran cultivados o eran también recogidos del campo sin cultivar?

Al considerar este asunto uno debe tener en cuenta, primero, que ni el trigo escandal ni los otros dos cereales se encuentran hoy en estado silvestre en ninguna parte cercana a Abu Hureyra; la zona es demasiado árida. Pero dadas las condiciones de mayor humedad y frescura del Mesolítico tardío, han podido, sin embargo, habitar allí en el pasado. Esto plantea otro problema: el hecho de que los granos de cereal recobrados de la aldea mesolítica de Abu Hureyra sean tipos morfológicamente silvestres ¿significa que no han podido ser cultivados? No necesariamente.

Hillman utilizó herbajes de espelta silvestre que crecen en las montañas de Monzur, al este de Turquía, para hacer experimentos con varios métodos de recolección y con esquemas de rotación de cultivos. Estos experimentos fueron realizados para comprobar si las presiones de la selección, favorables a la mutación a partir del estado silvestre, podrían operar en tales condiciones. Su conclusión es que en tanto que los métodos de cultivo adecuados para la simple labranza fuesen aplicados, la morfología de la espelta silvestre permanecería inalterada incluso durante largos períodos. Sólo después de que técnicas de cultivo no usuales se aplican durante muchos cientos de años, sostiene Hillman, aparecen las mutaciones del estado silvestre al doméstico.

Hay indicios que sugieren que los habitantes de Abu Hureyra practicaban justamente esos métodos simples. Por ejemplo, el cultivo de cereales, aun a pequeña escala, tenía que perturbar la vegetación natural de sus proximidades. Esto implica que los mismos hierbajos que hoy son característicos de las zonas cultivadas cercanas al yacimiento deberían estar presentes en las muestras de flotación de los niveles mesolíticos del mismo. Es posible, desde luego, que al ser pisoteado el suelo por los habitantes mesolíticos, o incluso por sus animales, se abriese una vía para estos mismos hierbajos. Como quiera que fuese, las muestras de flotación revelan que tres hierbajos actuales (*Atriplex*, uno de los tres miembros de la familia del galio que se crían en la estepa: *Alyssum*, un miembro de la familia de la mostaza que hoy es una planta de adorno, y *Lithospermum*, miembro de la familia de la borraja) crecían lo bastante cerca de Abu Hureyra para dejar sus semillas en la tierra excavada.

En conjunto, pues, parece altamente probable que los habitantes de la Abu Hureyra tardo-mesolítica cultivaran tres



ARTEFACTOS NEOLITICOS: puntas de flechas (*a*), cuchillas de hoces (*b*), escariadores (*c*) y buriles (*d*), hechos algunos de obsidiana importada, y muchos raspadores (*e-g*). Los objetos de hueso comprenden un gancho con un ojal (*h*), un anzuelo (*i*) y agujas (*j*). En las tumbas destacan las cuentas de "mariposa" (*k-n*), por lo regular de ágata o serpentina importadas. Aunque miden varios centímetros de superficie, su espesor se reduce a milímetros.

cereales (y quizá también algunas legumbres). Esta conclusión se halla confirmada por la semejanza entre los alrededores de Abu Hureyra, con mucha tierra potencialmente arable en su vecindad, y los alrededores de otros yacimientos del mesolítico tardío en Oriente: Dibsi Faraj Este y Mureybat, yacimientos del valle del Eúfrates descubiertos durante la última campaña de urgencia que vio las excavaciones de Abu Hureyra: Saaideh, en el Líbano, y Jericó, en Palestina. Sólo los yacimientos del Mesolítico temprano no están asociados con tierra cultivable.

Otro tipo de semejanza refuerza la conclusión de que una labranza simple estaba en marcha en Abu Hureyra. Por lo menos cuatro *tells* del Oriente que fueron sedes de florecientes comunidades agrícolas neolíticas han dado señales de habitación en los últimos tiempos mesolíticos. Tres de ellos se hallaban en la vecindad de tierra arable en potencia: Jericó, Mureybat y Abu Hureyra.

Pasando de los restos de plantas de Abu Hureyra a los restos de animales, Anthony J. Legge, de la Universidad de Londres, está analizando las dos toneladas de huesos de animales del yacimiento. Sus resultados muestran que los habitantes mataban especies de caza que variaban en tamaño desde el conejo hasta el onagro, el asno salvaje del Oriente. El animal predilecto como alimento era la gacela, pero los onagros, las ovejas y las cabras también eran muertos con asiduidad. La presencia, entre los huesos, de conchas de moluscos de agua dulce y de espinas de pescado indica que la población mesolítica de Abu Hureyra explotaba otra fuente de alimento: el Eúfrates.

En síntesis, los datos que poseemos indican que en tiempos del Mesolítico tardío los habitantes de Abu Hureyra estaban ya cultivando cereales (sus coetáneos en otros yacimientos del Oriente probablemente lo hacían así también). Al mismo tiempo consumían muchos alimentos vegetales y animales recogidos del mundo silvestre. Así pues, la tradición mesolítica básica de cazar y recolectar se había combinado aquí (y hasta cierto punto en otros lugares) con la agricultura rudimentaria. La ocasión para esta práctica estaba presente en forma de tierra potencialmente arable. El motivo pudo muy bien haber sido proporcionar la subsistencia necesaria para un nuevo género de vida, a saber, el establecimiento de una población incrementada en aldeas de cierto tamaño.

En el curso del milenio noveno a.C. la aldea mesolítica de Abu Hureyra fue abandonada por razones que ignoramos. El lugar estuvo desocupado durante unos mil años antes de que una nueva población apareciese en época del Neolítico arcaico. Por entonces, la agricultura era ampliamente practicada tanto aquí como en el resto del suroeste de Asia y los poblados sedentarios no constituían una novedad. Sin embargo, una vez que Abu Hureyra fue reocupada (hacia 7500 a.C.) pasó pronto a ser el mayor de los poblados del Neolítico arcaico del levante asiático. Miles de almas vivían en el gran *tell*.

Las consecuencias de que tanta gente se aglomerase en un espacio de menos de doce hectáreas eran muchas. En primer lugar, las relaciones sociales de los habitantes del *tell* tuvieron que ser con-

siderablemente más complejas que las del anterior período mesolítico; más aún, hubieron de ser más complejas que las relaciones sociales que se han atribuido a sociedades neolíticas tempranas. En segundo lugar, reglamentar las actividades de semejante poblado exigiría una forma efectiva de organización comunal. De ahí cabe deducir la aparición, aun en esta fecha tan temprana, de sistemas sociales y políticos que en forma más desarrollada habían de convertirse en rasgos esenciales de las primitivas ciudades-estado del suroeste de Asia, tanto en el Levante como en la misma Mesopotamia.

Considérese un exponente de la proyección al exterior de un yacimiento neolítico arcaico importante como lo era Abu Hureyra. Ese exponente es el comercio. Dado que Abu Hureyra era un poblado excepcionalmente grande en la segunda mitad del séptimo milenio, puede ser contemplado lógicamente como el poblado central de una región que abarcaba muchos poblados menores. Los materiales arqueológicos de Abu Hureyra sugieren que los componentes de este conjunto regional estaban en contacto no sólo unos con otros, sino también, y por primera vez, con otras regiones. Abu Hureyra recibía puntualmente el vidrio volcánico u obsidiana de varias fuentes de suministro en Turquía, junto con jadeíta, serpentina, ágata y malaquita. La esteatita le llegaba de los montes Zagros; la turquesa, del Sinaí; y las conchas de caurí, del Mediterráneo o del Mar Rojo. Por lo que respecta a qué ofrecían Abu Hureyra y las aldeas satélites a cambio de estas apetitosas materias primas, los registros arqueologicos permanecen mudos.

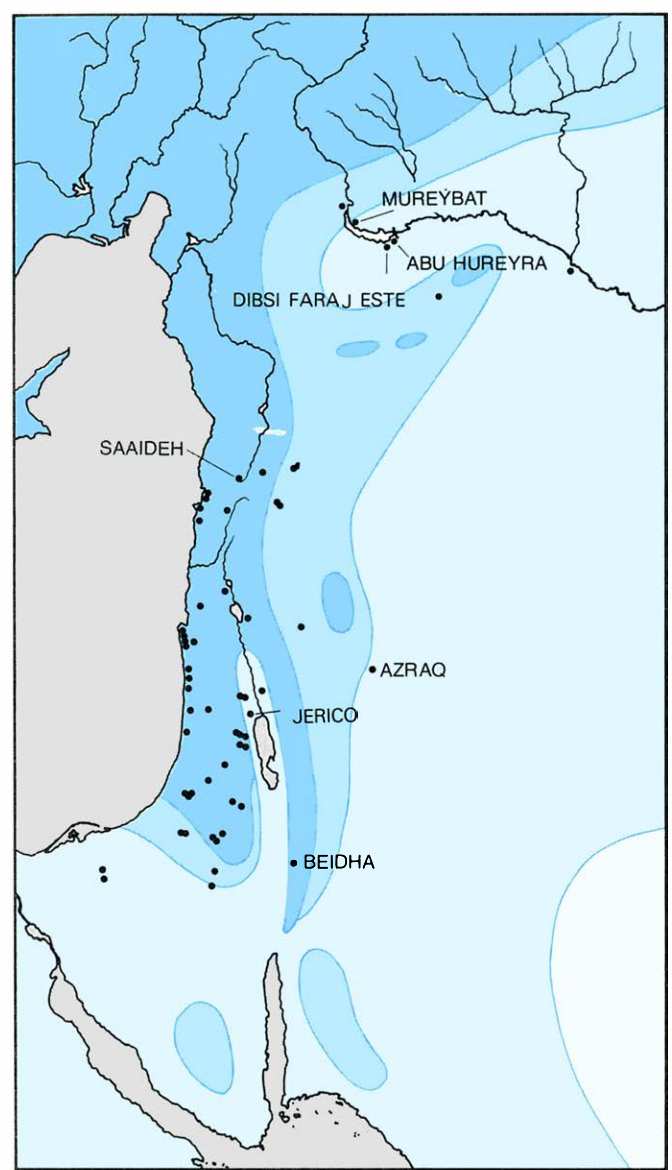
¿Qué factores determinaron el desusado crecimiento de Abu Hureyra como punto focal de una región? El factor individual más importante debió ser el desarrollo en el Cercano Oriente de un nuevo tipo de economía agrícola. En Abu Hureyra, como en algunos otros yacimientos del Neolítico inicial, esta nueva economía descansaba en el cultivo de cereales y de leguminosas. Mientras que es sólo probable que los predecesores mesolíticos de los aldeanos del Neolítico arcaico cultivasen la escanda, la cebada y las lentejas, es seguro que los aldeanos neolíticos lo hacían. Lo que quizá sea igualmente importante es que cultivaban también, además de la escanda, una nueva clase de trigo: la espelta. Aunque algunas otras plantas eran sembradas, cuatro de ellas –dos trigos, otro cereal y lentejas– constituían el fuerte de los cultivos.

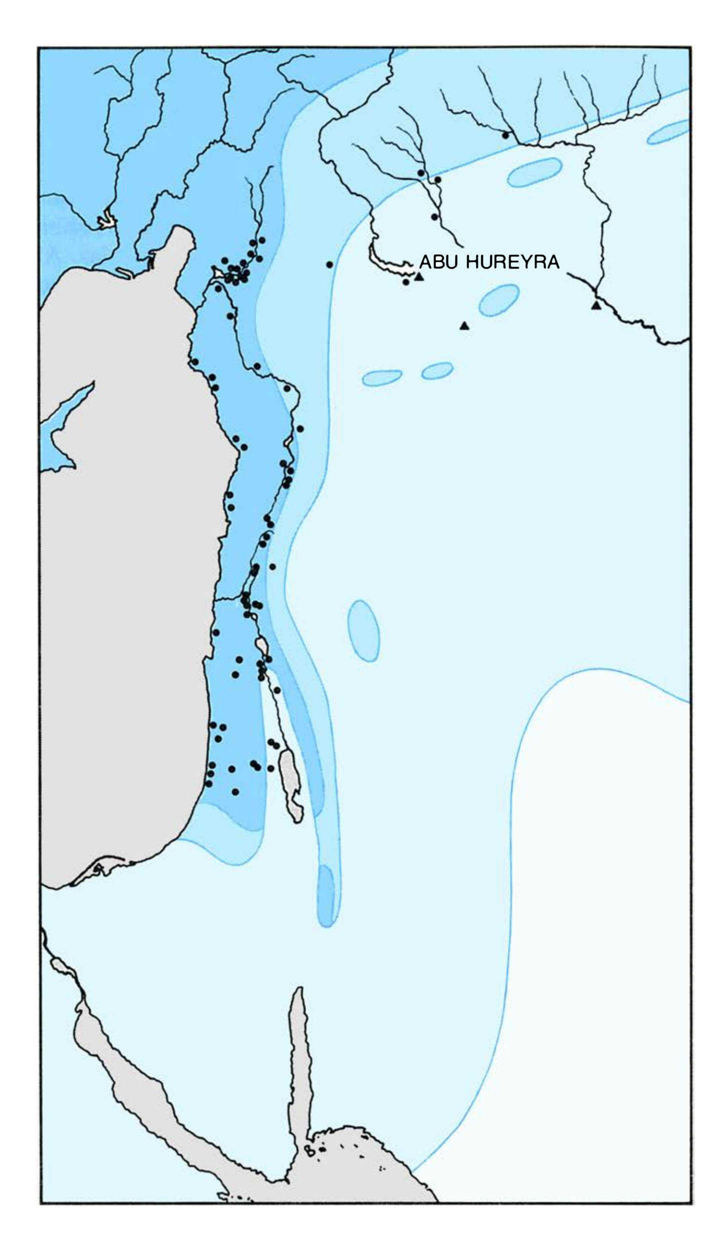
Otro factor en el crecimiento de Abu Hureyra fue su posición en la divisoria entre la llanura aluvial del Eúfrates y la estepa abierta hacia el sur. Esto permitía a la población humana explotar los recursos de los dos ambientes complementarios. Más aún, la gente de la Abu Hureyra neolítica debió aprovechar un accidente de la topografía local para poner en regadío las tierras bajas situadas al oeste del poblado.

En esta parte del valle del Eúfrates, hoy en día y en el pasado reciente, el regadío sólo ha sido posible con ayuda de bombas o con canales de muchos kilómetros de largo. Ninguno de estos métodos estaba al alcance de la población neolítica arcaica. Un poco al oeste de Abu Hureya, sin embargo, hay un valle pequeño, el Wadi Hibna. Hoy en día está seco, excepto después de un aguacero, pero en las circunstancias climáticas, más húmedas, de la época del Neolítico arcaico, probablemente llevaba una corriente permanente que podría ser desviada con facilidad. Con ello las gentes de Abu Hureyra podían poner en regadío tanto el fondo del *wadi* como la llanura aluvial al oeste del poblado, incrementando así considerablemente los productos de las cosechas y satisfaciendo las demandas de una población en crecimiento.

SECUENCIA EN EL LEVANTE ASIATICO	C-14 AÑOS A. C. (SIN CALIBRAR)		OCUPACION DE TELL ABU HUREYRA
	19,000		
MESOLITICO	18,000		
TEMPRANO			
	11,000		
	10,000		
MESOLITICO		DECIMO MILENIO	?
TARDIO	9,000		
		NOVENO MILENIO	?
	8,000		
NEOLITICO		OCTAVO MILENIO	
ARCAICO	7,000		
		SEPTIMO MILENIO	
	6,000		
		SEXTO MILENIO	
NEOLITICO			
	5,000		
DESARROLLADO			
		QUINTO MILENIO	
	4,000		

TRES FASES DE LA OCUPACION de Abu Hureyra (*barras en color*). El asentamiento empezó durante el período mesolítico tardío del levante asiático, hace unos 11.500 años. Los aldeanos mesolíticos abandonaron el lugar, por razones ignoradas, hace unos 10.500 años. Unos mil años después, Abu Hureyra volvió a ser ocupada por un grupo de granjeros cuya estrategia de subsistencia era característica del período neolítico arcaico. La nueva estación llegó pronto a ser la mayor de su clase en el Levante. La transición del Neolítico arcaico al avanzado se produjo unos 1500 años más tarde. El poblado fue abandonado no mucho después, como consecuencia de la disminución de las lluvias.





EL CAMBIO CLIMATICO del Levante entre el noveno milenio (*izquierda*) y el sexto milenio a.C. (*derecha*) es evidente en el desplazamiento de las manchas de vegetación. La abundante pluviosidad de la época mesolítica tardía provocó la expansión hacia el este de la zona de bosque mediterráneo, con más de 300 milímetros de lluvia al año (*color obscuro*) y de la zona intermedia de bosque abierto y de estepa a continuación, con lluvia de 300 a 150 milímetros al año (*colores más claros*). Mientras que la mayoría de los 70 yacimientos mesolíticos tardíos conocidos (*puntos negros*) estaban en la zona de máxima precipitación o en la zona intermedia, algunos se encontraban en rincones especialmente favorables de la estepa. Por ejemplo, Abu Hureyra y Dibsi Faraj Este se hallaban junto al Eúfrates, y Azraq estaba en un oasis. Tres milenios más tarde, el bosque compacto, el bosque abierto y la estepa retrocedieron hacia el oeste conforme las precipitaciones disminuían y el desierto (*color más claro*) avanzaba. Incluso las plazas bien situadas de la estepa (*triángulos*) fueron abandonadas entonces. De los 80 yacimientos conocidos del Neolítico del sexto milenio, la mayoría estaban dentro del área más favorable para las granjas mixtas, la zona del bosque mediterráneo. La economía de granja mixta es aún la base de subsistencia.

En cuanto al sector animal de la economía neolítica arcaica, están presentes las tres especies que eran perseguidas como caza por los ocupantes mesolíticos del yacimiento: la gacela, la oveja y la cabra. Sin embargo, ahora las tres especies eran probablemente criadas como rebaños, en lugar de cazadas. Los restos óseos muestran que al principio eran sacrificadas muchas más gacelas que ovejas o cabras. Después disminuyó de repente el número de gacelas y aumentó proporcionalmente el de ovejas y cabras. En la época en que la fase desarrollada del Neolítico había reemplazado a la fase arcaica, a principios del sexto milenio a.C., los bueyes y los cerdos se habían sumado a los rebaños del poblado y la gacela prácticamente había desparecido del medio.

En los restantes siglos del Neolítico arcaico, sin embargo, los tres animales originarios eran la principal fuente de carne en Abu Hureyra, y la caza y la recolección de alimentos complementarios seguía formando parte de la estrategia de subsistencia. Plantas silvestres y animales salvajes contribuían a la variedad de la dieta, como lo hacían el pescado y el marisco.

¿Qué aspecto tenía la Abu Hureyra del período neolítico? Sus numerosas casas rectilíneas estaban construidas apretadamente, con apenas estrechos callejones y patinillos entre ellas. Cada casa, hecha de adobe, consistía en varias habitaciones pequeñas, comunicadas por puertas; algunas de éstas eran del tipo corriente, pero otras tenían el umbral muy alto. Muchas habitaciones tenían pisos de yeso enlucido, que eran pintados de negro y a veces tenían dibujos esquemáticos en rojo. Las paredes de las casas estaban cubiertas de un enlucido de barro; algunas recibían una mano de cal. Parece probable que cada casa estuviese ocupada por una familia.

Los artefactos hechos por este grupo abarcaban útiles de sílex y, menos comúnmente, de obsidiana, con una gran

variedad de nuevas formas. La mayoría estaban hechos de lascas, fabricadas arrancando grandes trozos a un "núcleo". Los artefactos más corrientes de lasca son puntas de flecha, cuchillos, raspadores de punta (utensilios que actuaban con un extremo más que con un lado) y buriles (hojas talladas de modo que conservaran una afilada proyección en ángulo para grabar). Un número menor de lascas era convertido en escariadores y filos de hoces. Las puntas de flecha deben de haber estado destinadas preferentemente a la caza; pero encontramos una alojada en el pecho de un esqueleto humano enterrado en el poblado, lo que sugiere que el arco y la flecha también se empleaban para combatir.

Lawrence H. Keeley, de la Universidad de Illinois, y Emily Moss, de la Universidad de Londres, examinaron estos utensilios a grandes aumentos en busca de señales de desgaste que indicasen el empleo a que eran sometidos. Han podido demostrar que los raspadores de punta se usaban para curtir pieles, los buriles para grabar madera y hueso y los escariadores para hacer agujeros en esos mismos materiales.

La lista de los utensilios de piedra tallada la completan cuchillos y raspadores, hechos de grandes hojas, y martillos de piedra. Las hachas y los cinceles se hacían de piedra que no era tallada, sino pulimentada. Las trituradoras y piedras de moler utilizadas en la preparación de los cereales recibían su forma picándolas o raspándolas. Algunos cuencos de mejor calidad se hacían de piedra pulimentada; otros, más burdos, a costa de frotar. Aunque la mayoría de los yacimientos neolíticos contienen cerámica, Abu Hureyra y algunos otros yacimientos del suroeste de Asia no ofrecen cerámica alguna en esta fase de su ocupación. Una de las principales funciones de la cerámica es la de almacenar alimentos; pero estos poblados neolíticos, aparentemente, los almacenaban por lo menos de otra manera. Tenían grandes vasijas rectangulares hechas de grueso yeso blanco, tan pesadas que deben de haber estado fijadas permanentemente al lugar que se les asignaba en las habitaciones donde las hemos encontrado. Es de señalar que las proporciones de las varias clases de instrumentos de piedra difieren sensiblemente entre una y otra parte de Abu Hureyra, señal de que los habitantes del lugar en época neolítica arcaica hacían cosas distintas en las diferentes partes del poblado.

Los artefactos de hueso del Neolítico arcaico del yacimiento repiten el inventario mesolítico de agujas, punzones y espátulas. Un anzuelo de hueso y un gancho perforado de hueso, posiblemente empleado para colgar ropa, son adiciones neolíticas arcaicas. Los excavadores también descubrieron restos de cestos y de esteras. Estos deben de haber figurado entre los productos más corrientes de la artesanía del poblado; las cestas serían el complemento de las vasijas de yeso rectangulares para el almacenaje de alimento.

Así como el repertorio de objetos de adorno de los niveles mesolíticos del yacimiento era limitado, el número y variedad de tales artefactos en los niveles neolíticos arcaicos eran considerables. Hay cuentas y colgantes hechos de barro cocido, hueso y concha. Otros adornos están fabricados en una serie de piedras: simple caliza, varias piedras verdes, incluida la serpentina, y piedras semipreciosas tales como el ágata, la cornalina y la turquesa. Los más delicados de todos estos objetos, casi todos hallados en tumbas, pueden describirse como cuentas de mariposa. De forma trapezoidal o redondeada, están hechas por lo general de serpentina o de ágata. Las cuentas miden hasta cinco centímetros y medio en su dimensión mayor, pero habían sido rebajadas por frotamiento hasta no tener más que unos milímetros de espesor.

El uso de tales materiales, muchos de ellos importados, para la manufactura de artículos de adorno, sugiere que uno de los incentivos para el mantenimiento del comercio con comunidades distantes era la vanidad personal. Esta sugerencia se halla robustecida por el descubrimiento del uso último de otro material importado: la malaquita. Una caracola pequeña que desenterramos contenía restos de malaquita en polvo, un popular cosmético en épocas más recientes.

La presencia de cuentas hechas de barro cocido, entre estos adornos, es también notable. Otros objetos hechos de barro se hallaron en los niveles neolíticos arcaicos del yacimiento, especialmente figuritas de forma humana y animal. Estas figuritas, empero, no estaban cocidas. Sin embargo, si las cuentas de barro eran hechas en la localidad y no importadas, los habitantes de Abu Hureyra estarían familiarizados con dos propiedades básicas del barro que debieran haberles llevado al umbral de la manufactura de cerámica; la plasticidad y el endurecimiento al fuego.

El yacimiento neolítico arcaico alcanzó el máximo de su tamaño en el séptimo milenio a.C. Su crecimiento reflejaba el rápido aumento de pobla-

ARTES DECORATIVAS de la Abu Hureyra neolítica. Se manifiestan en una concha (*izquierda*) que contenía malaquita en polvo, un cosmético importado, y una gacela en miniatura hecha en granito.

ción experimentado por el Levante en conjunto, conforme se difundía la nueva economía agrícola. La ocupación de Abu Hureyra continuó después del comienzo del período neolítico avanzado, en el sexto milenio, pero la aldea se redujo a poco más de la mitad de su tamaño anterior. Como el poblado precedente, la aldea del Neolítico avanzado también estaba hecha de adobe. Poco se conoce de su aspecto por haber sido casi destruida por miles de años de exposición a los elementos. Los edificios parecen haber cambiado muy poco respecto a los del Neolítico arcaico, pero sus habitantes dieron en cavar numerosos fosos entre aquéllos. Los usaron algún tiempo como hogares y luego los llenaron de escombros.

Los artefactos hechos por los últimos habitantes de Abu Hureyra diferían poco de los de sus predecesores del Neolítico arcaico. Una excepción es que ahora aparece la cerámica. Incluso esta innovación parece haber sido de escala modesta; muy pocos tiestos se encontraron en esta parte de la excavación. Al final, alrededor del 5500 a.C., o quizás un poco antes, el lugar quedó abandonado. Y permaneció sin habitar hasta que nuestro grupo llegó para excavarlo.

¿Por qué fue Abu Hureyra abandonada por sus habitantes? Los resultados de la excavación indican que, durante la larga vida del poblado, se modificó su economía agrícola para hacerse cada vez más productiva. Podríamos buscar la razón del abandono a nivel puramente local, si no fuese por el hecho de que a través de todo el Levante asiático otros yacimientos del borde de la zona esteparia fueron abandonados también hacia el mismo momento.

Yo propongo dos razones para estos abandonos. La primera es que desde finales del séptimo milenio el ambiente de conjunto del Levante se iba haciendo menos atractivo para una población de colonos. La temperatura había estado subiendo desde la retirada de los hielos del Pleistoceno, pero la pluviosidad estaba ahora disminuyendo. El cambio provocó una reducción gradual de la zona del bosque mediterráneo. La estepa se extendió por las áreas boscosas, y a su vez el desierto comenzó a superponerse a la estepa. Así, mientras el Levante se tornaba más árido, el esquema del poblamiento a lo largo del borde de la estepa se veía también afectado.

La segunda causa de los abandonos, a mi modo de ver, radica en que el crecimiento de población durante el Neolítico arcaico impuso una carga demasiado pesada en la delicada balanza de la economía general. Como hemos visto, aquella economía tenía un carácter dual. Los productos de la agricultura y de la ganadería satisfacían muchas de las necesidades de la población, pero al mismo tiempo, la explotación de los recursos silvestres aportaba su propia e importante contribución. Los restos de plantas de Abu Hureyra indican que la vegetación natural (y por tanto, el terreno vegetal) se fue degradando progresivamente durante la vida del poblado. Es evidente que los habitantes estaban apacentando demasiado ganado y explotando en exceso la vegetación natural como fuente de alimento y de combustible. Parece probable que redujesen también seriamente la fauna de caza en un momento en que el cambio de clima estaba acelerando ese mismo proceso.

A comienzos del sexto milenio, posiblemente en respuesta a las presiones duales de un medio ambiente alterado y de un equilibrio ecológico trastornado, los habitantes de los poblados del Levante, incluido el de Abu Hureyra, dejaron de confiar en la caza y en la recolección como suplemento importante de la economía agraria y pasaron a depender casi exclusivamente del producto derivado de la agricultura y ganadería mixta (granja mixta). En sus siembras se concentraron en los granos cereales más productivos y en aquellas leguminosas que, como la lenteja, eran una abundante fuente de judías y guisantes. Sus rebaños incluían ahora ganado vacuno y cerdos, además de las ovejas y cabras de toda la vida.

Este tipo de economía de granja mixta se adaptaba mucho mejor a aquellas partes del Levante donde el bosque mediterráneo aún existía, es decir, a la costa, bien abastecida de agua, y a las llanuras del norte. Los habitantes de aquellas aldeas limítrofes de la estepa, florecientes un día, abandonaron sus exhaustos poblados y se trasladaron al norte y al oeste para erigir nuevos hogares en ambientes más prometedores. El hecho de que en conjunto la población del Levante en el sexto milenio a.C. siguiese siendo casi igual a como era a finales del séptimo milenio, caracterizado por su expansionismo, indica que el reajuste tuvo éxito, cualesquiera que fuesen los costos de la operación. Aunque el período neolítico estaba ya muy adelantado en el Levante antes de que se desarrollase una economía de granja mixta, la adopción definitiva de esta economía, a principios del sexto milenio, fue tan satisfactoria que ha seguido siendo la base de la existencia humana en aquella zona hasta la época moderna.

Ciencia y sociedad

Percepción visual del movimiento en profundidad

Los juicios visuales de una persona atareada en una calle de gran ajetreo son una verdadera hazaña. Toda la información visual que entra en el cerebro proviene del par de imágenes bidimensionales formadas cuando las lentes del ojo hacen converger los rayos incidentes sobre la retina. En base a estas dos imágenes bidimensionales, no demasiado perfectas, el peatón tiene que juzgar la posición y la dirección de los automóviles, las de la demás gente, así como las suyas propias respecto del mundo circundante tridimensional y en continuo movimiento. ¿Cómo procesa la información contenida en estas imágenes el sistema visual humano para permitir la precisa coordinación entre la vista y las extremidades que hace posible la correcta maniobra por entre el tránsito?

A pesar de los impresionantes adelantos registrados en el estudio de la percepción visual en los últimos 15 años, sólo se comprende todavía de forma muy parcial cómo se lleva a cabo la coordinación entre los ojos y las extremidades en el espacio tridimensional cuando tanto el sujeto como los objetos de su alrededor están en movimiento. David Regan, Kenneth Beverley y Max Cynader acaban de publicar en *Scientific American* un modelo explicativo que exponemos aquí resumido.

Ya en tiempos de Leonardo da Vinci se sabía que cada uno de los ojos del hombre, cuyas pupilas distan aproximadamente 6,5 centímetros, reciben imágenes ligeramente distintas del mundo visual y que, en consecuencia, el dibujo o la pintura bidimensional de un objeto sólido no puede ser siempre tomado como su imagen real. Sin embargo, no fue hasta el siglo XIX cuando se empezó a comprender de forma más clara cómo utiliza el cerebro las diferencias geométricas entre las imágenes retinianas derecha e izquierda de un objeto sólido. En 1838, el físico británico Sir Charles Wheatstone demostró, en la Royal Society de Londres, la importancia de estas diferencias geométricas, llamada disimilitud (disparity) binocular, para la localización precisa de los objetos en el espacio visual. Wheatstone dibujó algunos objetos basándose en cómo los veía cada ojo por separado y los presentó, luego, a los ojos correspondientes a través de un aparato que denominaba estereoscopio: un bonito instrumento de metal y madera pulida no muy diferente de los modernos visores estereoscópicos.

La mayoría de las personas que miraban por el estereoscopio fundían el par de imágenes en una sola que parecía alargarse en profundidad. Así pues, Wheatstone había demostrado que una diferencia geométrica entre la imagen retiniana derecha y la izquierda podía ser suficiente para producir la sensación de profundidad. Al año de la demostración de Wheatstone en la Royal Society, la fotografía ya era una realidad, así que pudo, utilizando pares de daguerrotipos tomados desde posiciones ligeramente distintas, mostrar que su descubrimiento no sólo era válido con los dibujos a mano alzada, sino también con exactas representaciones bidimensionales de los objetos. En los últimos 20 años, Bela Julesz, de los Bell Laboratories, ha continuado los trabajos de Wheatstone consiguiendo descifrar cómo el sistema visual compara las dos imágenes retinianas punto por punto para así responder al objeto como un todo.

El canal visual de la información de profundidad y del carácter sólido responde apreciablemente a los estímulos de disimilitud binocular. Horace B. Barlow, Colin Blakemore y John D. Pettigrew, de la Universidad de California en Berkeley, P. O. Bishop, de la Universidad Nacional de Australia, y Tosaku Nikara, de la Universidad de Sidney, descubrieron, en 1968, las bases neurales de este canal estereoscópico de posición. Registrando la actividad de células nerviosas de la corteza visual de gato con microelectrodos, descubrieron unas clases de neuronas que respondían particularmente a estímulos formados por pares de trazos con cierta disimilitud binocular. Estas neuronas podían ser altamente selectivas, pues sólo presentaban actividad cuando la disimilitud entre los dos trazos del mismo par se mantenía dentro de un estrecho margen.

Dando por sentada la existencia de esta clase de neuronas sensibles a la disimilitud, habría que plantearse el papel que juegan en la percepción del movimiento en profundidad así como en la percepción de la posición en profundidad.

En 1973, Regan y Beverley sugirieron que la trayectoria de un objeto moviéndose en profundidad podía ser determinada a partir de las velocidades relativas de las imágenes retiniana derecha e izquierda del objeto. Para contrastar esta hipótesis, realizamos unos experimentos de adaptación selectiva a estímulos de movimiento en profundidad generados en una versión moderna del estereoscopio de Wheatstone. Patrones oscilatorios independientes generados en un par de tubos de rayos catódicos fueron presentados por separado a cada uno de los ojos de los sujetos experimentales. La velocidad y la dirección del desplazamiento de los patrones (unas veces campos de puntos, otras trazos oscilando horizontalmente) eran controladas y la fusión de las imágenes aparecía en una pantalla frente al sujeto. Al mover los dos patrones en direcciones opuestas, la imagen resultante se parecía a una película tridimensional: los sujetos veían cierta expansión de los puntos o de los trazos que parecían moverse hacia adelante y hacia atrás en profundidad.

La sensación de movimiento en profundidad era considerable al principio de la presentación, pero disminuía progresivamente hasta que a los 10 minutos desaparecía. En otras palabras, habíamos creado unas condiciones de ceguera temporal al movimiento en profundidad. Estos resultados presuponían la existencia de un canal estereoscópico de desplazamiento, pues los movimientos hacia los lados podían ser percibidos tanto antes como después del período de adaptación. El canal de la percepción del movimiento en profundidad parecía ser más o menos independiente de los canales ya conocidos de percepción binocular del movimiento transversal.

Nos preocupaba también averiguar si los notables efectos de la adaptación estereoscópica al movimiento eran específicos de la dirección del desplazamiento del estímulo de adaptación. Por ejemplo, en sujetos adaptados a oscilaciones a lo largo de una recta que pasaba por la

izquierda de su nariz, se observaba una reducción de la sensibilidad visual para todas las trayectorias que pasasen a la izquierda de su nariz, pero no para aquellas otras que pasasen a la derecha. Parecían existir, por lo menos, dos canales estereoscópicos distintos y específicos del movimiento.

Estas observaciones sugerían que los canales eran selectivamente sensibles a la razón de las velocidades de las imágenes de las retinas izquierda y derecha, mientras que eran poco sensibles a la velocidad absoluta. Para determinar cuántos canales de este tipo existían en el sistema visual, trabajamos con sujetos adaptados a 13 direcciones distintas de desplazamiento en profundidad. Los resultados obtenidos parecían demostrar la existencia de sólo cuatro canales, cada uno de ellos funcional en ambos sentidos de una misma dirección. Estos canales diferían notablemente en el grado de selectividad direccional. Los cuatro canales sensibles al movimiento en alguna dirección que pase entre los ojos son altamente selectivos, respondiendo tan sólo a trayectorias comprendidas en un ángulo de 1,5 grados. Así pues, estos cuatro canales abarcan sólo 6 grados del campo visual. Los canales de percepción de movimiento cuya trayectoria no pasa por la cabeza responden a una gama mucho más amplia de direcciones, respondiendo a los restantes 354 grados del campo visual.

La idea de que la información relativa a la posición en profundidad y la relativa al movimiento en profundidad son procesadas por canales distintos se basa en observaciones distintas realizadas por Regan y por Whitman Richards, del MIT. Richards ha desarrollado una técnica, la estereoperimetría, para medir la distribución de la sensibilidad a la profundidad en las distintas regiones del campo visual de los sujetos. Esta técnica ha permitido la identificación de sujetos cuyo campo visual incluye áreas ciegas para el movimiento en profundidad, aunque normales para la sensibilidad a la posición en profundidad. Más recientemente, Richards ha podido encontrar individuos con el defecto inverso.

En la búsqueda del substrato fisiológico de estos canales y particularmente de las neuronas que responden selectivamente al desplazamiento estereoscópico en profundidad, Regan y Cynader registraron, con microelectrodos, los patrones de descargas de algunas neuronas del área 18 de la corteza visual de gatos anestesiados. El estímulo era, en este caso, un par de trazos que oscilaban independientemente de un lado a otro y que se presentaban por separado a uno y otro ojo del animal, mientras un experimentador intentaba que los trazos aparecieran como una sola imagen moviéndose hacia adelante y hacia atrás. Regulando las velocidades relativas de los dos trazos era posible variar la dirección del movimiento en profundidad (en un solo plano horizontal), de suerte que se podía registrar las respuestas de las neuronas al movimiento en cualquiera de las direcciones dentro de los 360 grados.

Aproximadamente la tercera parte de las neuronas cuya actividad se registró mostró ser sensible a la dirección del movimiento en profundidad. Estas células otorgaban mayor importancia a la información referente al movimiento que a la de la posición. Es importante considerar aquí que, en este experimento, la dirección del movimiento en profundidad no podía ser evaluada con un solo ojo (de hecho, el movimiento en profundidad no puede ser siempre percibido con un solo ojo). La dirección del movimiento del estímulo solo podía ser evaluada por el cerebro del gato mediante la comparación de las señales eferentes de los ojos derecho e izquierdo. Así pues, nuestros resultados reflejaban la interacción de las señales de los dos ojos.

Al objeto de medir esta interacción binocular registramos también la respuesta de las mismas neuronas a una amplia gama de direcciones del estímulo con gatos a los que se cubría primero el ojo izquierdo y luego el derecho. Para un tipo de neuronas, la suma de la actividad en condiciones de percepción con uno y otro ojo era siempre mayor que la respuesta en condiciones binoculares, lo que demuestra que las señales de los dos ojos se inhiben mutuamente. La intensidad de esta inhibición variaba en función de la dirección del movimiento en profundidad del estímulo. Descubrimos también unas neuronas que conseguían una selectividad direccional parecida, gracias a un mecanismo neural distinto. Para esta clase de neuronas, la interacción binocular, en vez de la supresión de la actividad en respuesta al movimiento hacia los lados, facilitaba la actividad en la dirección preferida, de tal manera que la respuesta a la estimulación binocular era mucho mayor que la suma de las respuestas a estimulación monocular.

Varias de las muchas neuronas de la corteza visual de gato cuya actividad se había registrado parecían corresponderse con los distintos canales estereoscópicos de movimiento descritos en nuestros experimentos psicofísicos en sujetos humanos. De estas neuronas,

algunas eran específicas para el movimiento dirigido hacia la cabeza mientras que otras lo eran al alejamiento del objeto. Otras neuronas respondían al movimiento en un margen mucho más amplio de direcciones (a menudo más de 90 grados), alejadas de la cabeza. Estos resultados pertenecientes al del área 18 de la corteza visual de gato se han visto recientemente confirmados para el área 17 de la corteza visual de monos por Gian F. Poggio y William H. Talbot, de la Johns Hopkins University.

Nos interesaba también saber si, comparadas con las respuestas de neuronas sensibles a disimilitud estática, las respuestas de las neuronas sensibles al movimiento estereoscópico eran independientes de la distancia entre el estímulo y la cabeza del animal. Habíamos observado que la sensación de movimiento en profundidad puede ser producida tan sólo por la variación en el tamaño del estímulo o tan sólo por la percepción estereoscópica del movimiento del estímulo, y nos preguntábamos si había alguna diferencia cualitativa en la percepción del movimiento en profundidad a partir de uno u otro tipo de información visual. Empezamos el estudio de la cuestión provocando un postefecto del cambio de tamaño. Los sujetos experimentales fueron adaptados a un estímulo rectangular en expansión de tal manera que al presentarles posteriormente un rectángulo de lados inmóviles les parecía verlo alejarse en profundidad. Intentamos luego la extinción de este postefecto no con un estímulo de tamaño variable sino con un estímulo de movimiento estereoscópico: rectángulos idénticos que se alejaban el uno del otro horizontalmente fueron presentados independientemente a cada uno de los ojos, de manera que la fusión de las imágenes provocara la sensación de movimiento en profundidad hacia la cabeza del sujeto. En los individuos de sensibilidad estereoscópica normal al movimiento, el postefecto de cambio de tamaño podía eliminarse fácilmente así.

Todo ello indicaba que la modificación del tamaño y el movimiento estereoscópico producían señales visuales que convergían en el mismo centro del movimiento en profundidad del sistema perceptivo visual. Se cosecharon nuevas pruebas en favor de este modelo teórico al generar un estímulo cuadrado en el que las velocidades relativas de las imágenes retinianas izquierda y derecha señalaban movimiento hacia la cabeza, mientras que la variación en el tamaño indicaba alejamiento de la cabeza. En este caso, la información facilitada por la velocidad relativa era equivalente a la de la disimilitud binocular. Regulando la razón de la disimilitud fue posible eliminar la sensación de movimiento en profundidad.

En el mundo real, la modificación de tamaño y la variación en la disimilitud no son casi nunca estimulaciones contrapuestas como en nuestros experimentos. En nuestro diseño experimental es posible, sin embargo, comparar directamente las dos señales de entrada; esta comparación se manifiesta útil en la detección del movimiento real de los objetos por el sistema visual. La razón entre el cambio de tamaño y la variación en la disimilitud de un objeto en movimiento es constante y depende de la anchura del objeto y de la separación entre los ojos del observador, pero es, sorprendentemente, independiente de la distancia a la que es visto. Al normalizar nuestros

datos respecto de esta razón, observamos que la velocidad en profundidad de un objeto y el tiempo que dura su observación ejercen un efecto considerable sobre la eficacia relativa de la modificación del tamaño y de la variación de la disimilitud como estímulos para la percepción del movimiento en profundidad.

Música celestial

Puede que a partir de ahora haya que utilizar con más cuidado esta expresión habitual mediante la que solemos calificar aquello que, pareciendo verosímil, no podrá ser nunca experimentado por nosotros. Por lo pronto, cualquiera puede oír ya una determinada música celestial, la contenida en el disco de larga duración Kepler LP 1571.

Johannes Kepler, uno de los grandes astrónomos de todos los tiempos y el que estableció las leyes definitivas del movimiento de los planetas, tuvo una especial dedicación a la idea de que la estructura y regularidad de los movimientos celestes eran como una gran partitura musical. En el contexto de un saber humanista, y concretamente neoplatónico, la relación entre las diversas ramas del conocimiento resultaba mucho más natural de lo que nos parece hoy. Por otro lado, téngase en cuenta que la teoría musical era la disciplina más avanzada en el registro y regulación de tiempos y ritmos. Parece que Galileo cronometró muchos de sus experimentos sobre la caída de los sólidos mediante ritmos musicales. En todo caso, Kepler publicó en 1619 una obra titulada *Harmonices Mundi libri V*, continuación del *Mysterium Cosmographicum*, en el que culminaba su teoría sobre la armonía del universo, idea que le tenía arrebatada su mente desde su juventud; allí descubre la tercera ley del movimiento de los planetas y, entre otras cosas, expresa la velocidad angular variable de cada uno de los planetas conocidos, en su órbita alrededor del sol, mediante series de notas musicales. Estas notas se repiten continuamente en cada revolución orbital y, todas juntas, forman una especie de concierto que es la música de las esferas celestiales. Como Mercurio da 122 vueltas al Sol mientras Saturno da una, aquél interpreta su tema 122 veces más rápido que éste.

Esta partitura celestial, cuyo sonido Kepler sólo pudo probablemente imaginarse, ha sido interpretada ahora en la Tierra, más de 350 años después, gracias a dos cacharros tecnológicos contemporáneos, la computadora y el sintetizador de sonidos, y a los esfuerzos de Willie Ruff y John Rodgers, de la Universidad de Yale, y de Mark Rosenberg, de Princeton, según informa la revista *Sky and Telescope* en su número de agosto.

La duración del disco es de 40 minutos y abarca 264 años de movimientos planetarios, desde el nacimiento de Kepler el 27 de diciembre de 1571 hasta el mismo mes de 1835. Cada uno de los nueve planetas (pues se han añadido los tres más exteriores desconocidos para Kepler) tiene su sonido característico. Mercurio suena como una flauta dulce con una melodía rápidamente modulada. Como Venus tiene una órbita casi circular, su sonido varía muy poco y es una especie de zumbido. Júpiter y Saturno tienen a su cargo la percusión.

El crítico musical Harold C. Schonberg ha escrito en el *New York Times*: "El sonido se parece mucho a la música electrónica de vanguardia que se elaboró

en los años 50 y 60, gracias al uso frecuente de sonidos diafónicos, a la relación inflexiblemente lógica entre elementos abstractos y a las pautas métricas establecidas por el computador con inhumana exactitud".

Líneas duras

La radiación electromagnética emitida en los procesos celestes incluye no sólo ondas de radio, radiación infrarroja, visible, ultravioleta y rayos X, sino también rayos gamma. A pesar de que ya se conocen desde hace algún tiempo los procesos responsables de la emisión de rayos gamma, sólo recientemente ha sido posible comenzar a detectar líneas espectrales correspondientes a las ondas electromagnéticas "más duras", o de mayor energía, que provienen del espacio exterior. A tal fin se han diseñado detectores que se incorporarán en un globo o en un satélite y elevados hasta las capas altas de la atmósfera terrestre, que es muy opaca a los rayos gamma.

Se tienen esperanzas fundadas en que los estudios espectroscópicos de los rayos gamma proporcionen una información fundamental sobre ciertos fenómenos de altas energías que se producen en el Sol y en las demás estrellas, supernovas incluidas. Según una reseña aparecida en un reciente número de *Nature,* aunque esos estudios se encuentran todavía en una fase inicial, ya se han detectado líneas correspondientes a rayos gamma en las erupciones solares, en el centro de nuestra galaxia, en la galaxia Centaurus A, fuente asimismo de radioondas, y en un acontecimiento fugaz no identificado todavía en la dirección de la Nebulosa del Cangrejo.

En las erupciones solares es en donde se observaron por vez primera líneas correspondientes a rayos gamma de origen cósmico. Probablemente la fuente de estas líneas sea la interacción entre las partículas aceleradas por las erupciones en la cromosfera y en la parte interior de la corona. La línea más intensa, con una longitud de onda equivalente a una energía de 2223 millones de electronvolt (MeV), es el resultado de la captura de un neutrón por un núcleo de hidrógeno. Otras dos líneas, a 7632 y 7646 MeV, son el resultado de la captura de un neutrón por un núcleo de hierro. La segunda línea en cuanto a intensidad, situada a 0,511 MeV, se debe a la aniquilación mutua de un electrón y un positrón (o electrón con carga positiva).

M. Leventhal, de los Laboratorios Bell, y C. J. MacCallum y P. D. Stang, de los Laboratorios Sandia, han escrito en *Astrophysical Journal Letters* que la línea de 0,511 MeV ha sido observada también en el centro de nuestra galaxia. Esta línea, cuya intensidad es cinco veces mayor que las fluctuaciones del fondo, fue descubierta mediante un gran detector, incorporado en un globo sobre Alice Springs, en Australia, que estaba apuntando directamente hacia el centro de nuestra galaxia. Debido a que la resolución angular del detector era sólo de 15 grados, la línea no pudo atribuirse a ninguna fuente determinada de radiación. Si la fuente se encuentra exactamente en el centro de la galaxia, está emitiendo una cantidad formidable de energía: unos 10^{30} joule por segundo (un joule es la energía necesaria para elevar 10 centímetros un kilogramo). Esta cantidad de energía es varios miles de veces mayor que la energía total emitida cada segundo por el Sol.

Los positrones cuya aniquilación da lugar a la raya de 0,511 MeV pueden haber sido creados en las reacciones nucleares de las explosiones de las supernovas. La muerte por explosión de una estrella de gran masa arroja al medio interestelar núcleos de níquel, cobalto, sodio, titanio, hierro y aluminio. Al producirse la desintegración radiactiva de estos núcleos, se emiten positrones. Otra posible fuente de positrones es la magnetosfera de un púlsar, aunque las observaciones realizadas sobre un púlsar particular no revelan ninguna línea debida a rayos gamma.

Se han asociado también ciertas líneas de emisión de rayos gamma con varios episodios fugaces. Desde hace varios años se vienen detectando ráfagas intensas de rayos gamma, algunas con una duración de quizá 10 segundos, aunque todavía no se han detectado con la suficiente resolución como para revelar si los rayos gamma han sido emitidos en líneas o en un continuo. Sin embargo, otra clase de acontecimientos fugaces presentan toda la radiación observada en forma de líneas. Uno de ellos, con una duración aproximada de veinte minutos, fue hallado en la Nebulosa del Cangrejo por un detector incorporado en un globo el 10 de junio de 1974. La intensa emisión de rayos gamma caía en cuatro bandas de energía comparativamente estrechas. La radiación estaba desplazada hacia el rojo de una forma muy marcada, es decir, su longitud de onda se veía aumentada por la pérdida de energía por tener que vencer un campo gravitatorio muy intenso. Esto sugiere que la emisión fugaz en cuestión está asociada con un campo gravitatorio ultrafuerte, como el de una estrella de neutrones o de un agujero negro.

Corales y arrecifes coralinos

Los diminutos pólipos de los corales, que viven en simbiosis con algas fotosintéticas, construyen gigantescos arrecifes calizos que albergan más especies vegetales y animales que ningún otro ecosistema

Thomas F. Goreau, Nora I. Goreau y Thomas J. Goreau

La capacidad que el hombre tiene de alterar la superficie de la Tierra sólo tiene un rival entre los organismos biológicos, las colonias de los diminutos pólipos de los corales, que a lo largo de épocas geológicas enteras han dado origen por acreción a gigantescos arrecifes calizos. Los verdaderos corales arrecifales o madréporas tienen su distribución geográfica limitada a las aguas claras, cálidas e iluminadas por el sol de los océanos tropicales; se encuentran en las grandes regiones arrecifales del Indopacífico y del Atlántico occidental. Los arrecifes son importantes constructores de tierra firme en las zonas tropicales, al formar cadenas completas de islas y alterar, asimismo, la línea de costa de los continentes.

Existen tres tipos principales de arrecifes coralinos. Los arrecifes costeros crecen en aguas someras y bordean una costa, muy cerca de ella o separados por un estrecho brazo de agua. Los arrecifes barrera se disponen asimismo paralelamente a una costa, pero están más alejados de ella, son mayores y se disponen sin solución de continuidad a lo largo de distancias más prolongadas; el mejor conocido es el Gran Arrecife Barrera (Great Barrier Reef), en aguas de la costa nororiental de Australia, que forma un baluarte subacuático de más de 2000 kilómetros de longitud y alcanza 145 kilómetros de anchura y 120 metros de altura. Los atolones son anillos de islas de coral que rodean una laguna central; cientos de ellos se hallan esparcidos por el Pacífico Sur. Están constituidos por arrecifes de varios miles de metros de diámetro, y muchos de ellos están formados por antiguos conos volcánicos que se han visto sometidos a un hundimiento y en los que la tasa de crecimiento del coral ha igualado a la tasa de subsidencia. Esta explicación de los atolones se debe a Charles Darwin, que la propuso durante el viaje del *Beagle*, y fue confirmada en los años cincuenta por Harry S. Ladd y Joshua I. Tracey, del U. S. Geological Survey, cuando sus extensos programas de perforación en atolones del Pacífico toparon con rocas volcánicas a cientos de metros de profundidad.

Aunque las aguas oceánicas tropicales son pobres en nutrientes, pues poseen reducidas concentraciones de nitratos, amonio y fosfatos en disolución, los ambientes de los arrecifes coralinos se cuentan entre los de mayores tasas de fijación fotosintética de carbono, de fijación de nitrógeno y de deposición de caliza de cualquier ecosistema. El ecosistema arrecifal mantiene también, con probabilidad, mayor número de especies animales y vegetales que cualquier otro. La clave de esta productividad prodigiosa es la biología única de los corales, que desempeña un papel fundamental en la estructura, la ecología y el ciclo de los nutrientes de la comunidad arrecifal.

La biología de los corales

Al ser sésiles, durante mucho tiempo se pensó que los corales eran plantas. En sus *Metamorfosis*, Ovidio se refiere al coral como un organismo que es blando bajo el agua pero que se endurece al entrar en contacto con el aire. (Lo que realmente vio Ovidio fue la muerte del tejido vivo, que deja al descubierto el esqueleto duro.) En 1723 el naturalista Jean André Peyssonnel propuso a la Academia Francesa de Ciencias que los corales eran animales. Su hipótesis fue ridiculizada, humillación que indujo al sabio a abandonar su trabajo científico. Las investigaciones posteriores, naturalmente, se encargaron de darle la razón. Los corales pertenecen al extenso y variado tipo animal de los Celentéreos, que son animales pluricelulares simples. El nombre del tipo deriva de los términos griegos *koilos*, cavidad, y *enteron*, intestino, porque la principal cavidad del cuerpo de sus miembros es la cavidad digestiva.

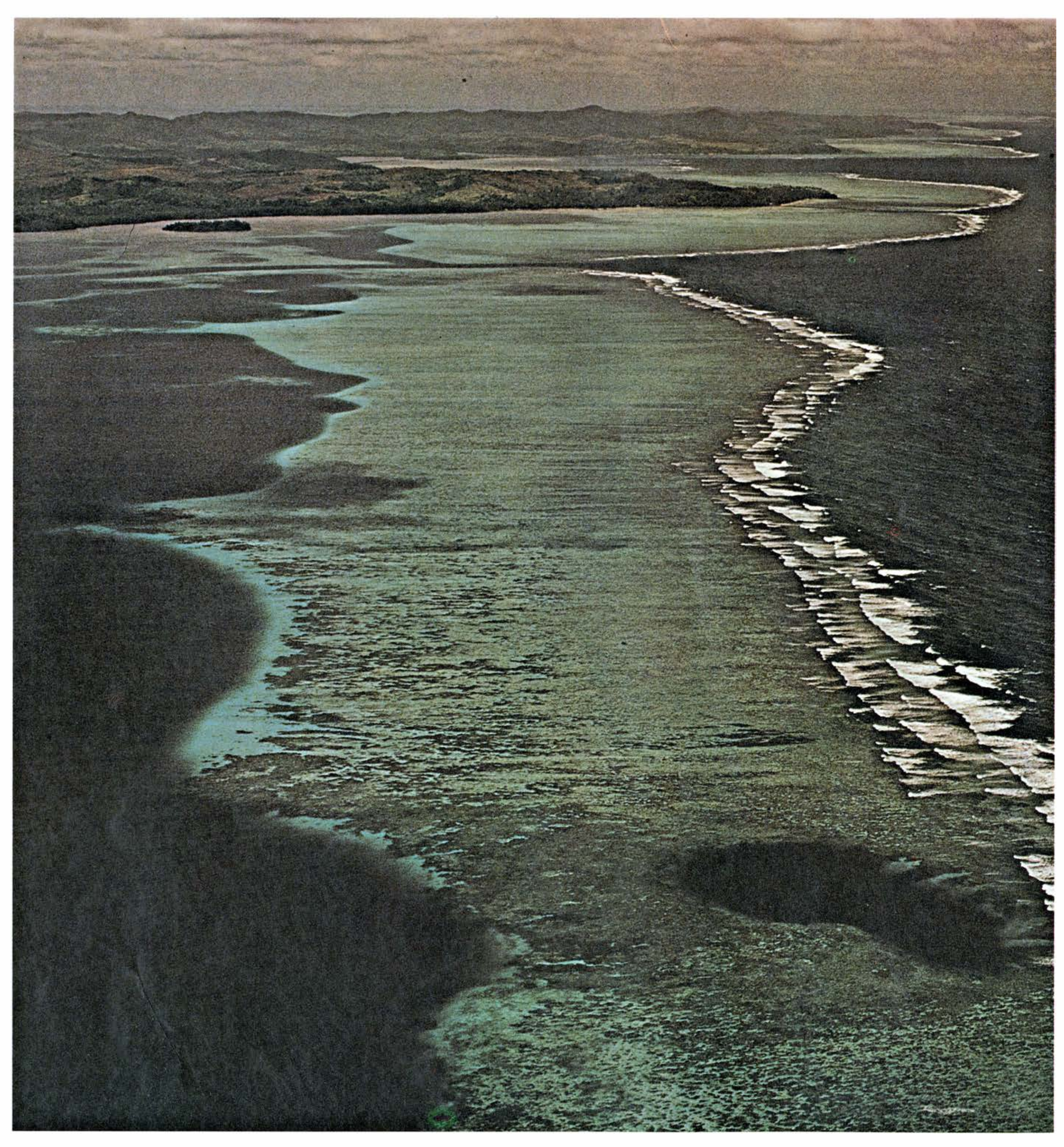
Los parientes más cercanos de los corales verdaderos son las anémonas de mar, a las que los corales se parecen en estructura somática básica y aspecto general. El blando pólipo de un coral está formado por tres capas de células y básicamente es un saco contráctil coronado por un anillo de seis tentáculos (o por un múltiplo de seis) que rodea una abertura parecida a una boca. Los tentáculos poseen células urticantes especializadas llamadas nematocistos, que disparan una flecha barbelada y una toxina que aturden a presas animales tales como crustáceos microscópicos. Desde la boca del pólipo, la corta faringe muscular desciende hasta la cavidad gástrica y se halla conectada a la pared del cuerpo mediante seis particiones (o un múltiplo de seis), lo que aumenta la superficie de la región digestiva. Los bordes libres de las particiones se hallan extendidas formando filamentos mesentéricos, que son tubos retorcidos que pueden evaginarse a través de la boca o de la pared del cuerpo.

El tamaño de los pólipos es muy variable, desde aproximadamente un milímetro de diámetro en algunas especies hasta más de 20 centímetros en otras. Cada pólipo puede dar origen a una gran colonia mediante división asexual o gemación. Los corales se reproducen también sexualmente, y producen larvas nadadoras que posteriormente se instalan y establecen nuevas colonias. La característica más sorprendente de las colonias de coral es su capacidad para formar un esqueleto calcáreo sólido. En muchos arrecifes abundan colonias coralinas individuales que pesan varios cientos de toneladas y que son suficientemente grandes como para ocupar una sala de estar. En muchas especies los pólipos se hallan en copas esqueléticas individuales: unos extienden sus tentáculos para alimentarse por la noche y otros los retiran parcialmente al interior de las copas durante el día. Cuando están contraídos, los pólipos pueden resistir la desecación

o el daño mecánico en marea baja, momento en el que algunas de las colonias pueden hallarse fuera del agua. Las copas esqueléticas se hallan constituidas por acúmulos de cristales de carbonato cálcico en forma de abanico, que se disponen en modelos que son típicos de cada especie de coral.

Una característica notable de todos los corales constructores de arrecifes es su simbiosis con las algas unicelulares denominadas zooxantelas. Los pólipos coralinos contienen gran número de estas algas en el interior de células que tapizan su intestino. Las zooxantelas son algas marinas pardoamarillas de la familia Dinofíceas, a la que pertenecen asimismo muchas algas dinoflageladas de vida libre. Las algas viven, realizan la fotosíntesis y se dividen en el interior de las células de su patrón u hospedador, la madrépora, y en esta simbiosis reside toda la productividad biológica del ecosistema del arrecife de coral.

Puesto que las zooxantelas de los corales formadores de arrecifes necesitan luz para realizar la fotosíntesis, estos corales viven únicamente en aguas oceánicas de menos de 100 metros de profundidad. Los corales requieren asimismo aguas cálidas (por encima de los 20 grados Celsius) y no toleran ni la salinidad reducida ni la turbidez elevada. En aquellos lugares en los que las colonias situadas a mayor profundidad se hallan a la sombra de las más



ARRECIFE DIMES, en aguas de las islas Palau, en el Pacífico occidental. La cresta somera del arrecife barrera está indicada por las olas que rompen sobre ella. Las islas Palau, del grupo de las Carolinas Occidentales, se hallan a 1060 millas al sudeste de Manila. (Foto aérea de Douglas Faulkner).



CORAL DE CUERNO DE CIERVO, especie que se encuentra de forma común en las aguas someras e iluminadas de los arrecifes coralinos tropicales. A lo largo de la cresta del arrecife, donde las madréporas tienen que soportar las fuerzas mecánicas del oleaje, las colonias de *Acropora cervicornis* son robustas y tienen ramas cortas. En las aguas resguardadas situadas detrás de la cresta del arrecife, las colonias son más altas y tienen ramas más largas y esbeltas, como las que se ven aquí. Los corales siempre crecen hacia la luz. Otros corales ramificados son los corales de cuerno de alce.

POLIPOS DE CORAL VIVOS, fotografiados a la puesta de sol, cuando emergen para alimentarse. Durante el día se retraen hacia el interior de sus copas esqueléticas, y así poder soportar la desecación si la colonia queda al descubierto en marea baja. Los pólipos, que se reproducen sexualmente y por división asexual, tapizan toda la superficie del esqueleto coralino. Se alimentan de animales del plancton, a los que aturden con los nematocistos.

someras, las colonias profundas maximizan su capacidad de captación de luz creciendo en ramificaciones como las ramas de los árboles del bosque. En aguas someras, donde la luz es abundante pero la tensión debida al oleaje es alta, las colonias depositan robustos esqueletos arborescentes; en aguas más profundas, donde la luz es escasa, las colonias forman estructuras horizontales tabulares en las que cada pólipo puede tener un número mayor de zooxantelas. En condiciones muy desfavorables, como las debidas a un período de oscuridad prolongado o a la inundación por agua dulce, para los pólipos coralinos ya no resulta ventajoso mantener a sus zooxantelas y las expulsan de sus tejidos. Puesto que la tasa de crecimiento esquelético de los corales depende de sus algas consortes, casi nunca se encuentran verdaderos corales constructores de arrecifes fuera de la zona de distribución de simbiosis estables.



DOS ESPECIES del coral *Agaricia* que crecen juntas difieren sorprendentemente en forma y tamaño. Una especie tiene frondes espiralados, mientras que la otra presenta placas en forma de teja. Estas complejas diferencias morfológicas son producidas por sutiles gradientes ambientales, como la disminución de la intensidad de la luz ambiente con la profundidad. Los corales jamaicanos se hallan a 43 m.

Hay especies de corales que no albergan zooxantelas; algunas de estas especies se encuentran en grietas situadas bajo las grandes estructuras erigidas por los corales constructores de arrecifes. La mayoría son corales solitarios en forma de copa, como *Astrangia*, que incrusta conchas y rocas hasta zonas tan septentrionales como Cape Cod. Estos corales pueden tolerar salinidades más reducidas, temperaturas más bajas y mayores profundidades: se han observado a 6000 metros en los grandes fondos marinos. Incluso en las aguas frías y profundas de los fiordos noruegos se hallan grandes bancos de *Lophophelia*, un coral colonial ramificado. Aunque estos corales no simbióticos se hallan distribuidos por todo el mundo, su tasa de crecimiento es mucho menor que la de sus parientes simbióticos, y no forman grandes arrecifes. En algunos casos aislados las colonias de estos corales contienen algas simbiontes, pero éstas no parecen contribuir de forma significativa a la nutrición de su patrón.

Las algas zooxantelas son almacenadas en el interior de cavidades individuales rodeadas por membranas, dentro de cada una de las células de la pared gástrica del pólipo coralino. El mecanismo de retroacción por el que el patrón regula el número de sus células algales no se ha determinado, pero no parece haber pruebas de que los corales "cultiven" y digieran sus algas. En cambio, los pólipos coralinos parecen controlar la población de zooxantelas mediante la expulsión de las algas más viejas y metabólicamente menos activas.

Robert K. Trench y sus colegas, de la Universidad de California en Santa Bárbara, han demostrado que cepas específicas de zooxantelas se hallan adaptadas a determinadas especies de corales. Algunas cepas pueden vivir bien en varios corales distintos, y otras madréporas no discriminan acerca del linaje de sus algas simbiontes. Los fascinantes problemas que plantea la selectividad simbiótica de los corales empiezan sólo a ser explorados, y los corales proporcionan un valioso sistema experimental para el estudio de las interacciones celulares en general.

Fisiología de la simbiosis de los corales

El estudio moderno de la fisiología de la simbiosis de los corales comenzó con una serie de elegantes experimentos realizados por C. M. Yonge durante la Great Barrier Reef Expedition de 1929. Yonge demostró que los corales simbióticos toman fosfatos y amonio del agua de mar circundante durante el día y devuelven estos nutrientes por la noche. Con el fin de estudiar con mayor detalle este fenómeno, dos de nosotros (Thomas F. Goreau y Nora I. Goreau) suministramos carbono en forma del isótopo radiactivo carbono 14 a corales arrecifales. Durante las horas de luz diurna las zooxantelas asimilaron el carbón marcado radiactivamente y lo fijaron fotosintéticamente en materia orgánica a una tasa que dependía de la intensidad de la luz. Parte de esta materia orgánica "pasaba" entonces desde el alga al coral hospedador. Estudios posteriores realizados por Trench y Leonard Muscatine, de la Universidad de California en Los Angeles, y por David Smith, de la Universidad de Oxford, demostraron que los compuestos que son transmitidos incluyen nutrientes simples como glicerol, glucosa y aminoácidos. Estos compuestos son utilizados por los pólipos coralinos en rutas metabólicas productoras de energía, o como bloques básicos para la elaboración de proteínas, grasas y carbohidratos.

Hace tiempo que se sabe que las tasas de las reacciones metabólicas se hallan limitadas estrictamente por la tasa a la que los productos de desecho son eliminados del entorno inmediato. En los animales superiores esta tarea corre a cargo de sistemas circulatorios y excretores especializados. Estos sistemas faltan en los Celentéreos, anatómicamente simples, que cuentan sobre todo con el lento proceso de difusión para eliminar los productos de desecho inorgánicos y solubles como el dióxido de carbono, fosfatos, nitratos, sulfatos y el amonio. Sin embargo, las zooxantelas necesitan para la fotosíntesis las mismas sustancias de las que el pólipo coralino debe desembarazarse, y se cree que las algas las toman activamente de su patrón.

De modo que los requerimientos fotosintéticos de las zooxantelas consiguen el ciclado de los productos de desecho del coral en nueva materia

orgánica. Durante las horas de luz diurna las algas simbiontes producen más oxígeno del que el pólipo coralino puede utilizar para su respiración, y parte del dióxido de carbono producido por el proceso respiratorio es vuelto a fijar por las algas en nueva materia orgánica. Con el fin de estimar la eficiencia del ciclado del carbono interno en las madréporas, uno de nosotros (Thomas J. Goreau) determinó la abundancia en el tejido y el esqueleto coralinos de carbono 13, un isótopo natural raro pero no radiactivo, con respecto a la abundancia del isótopo natural común, el carbono 12.

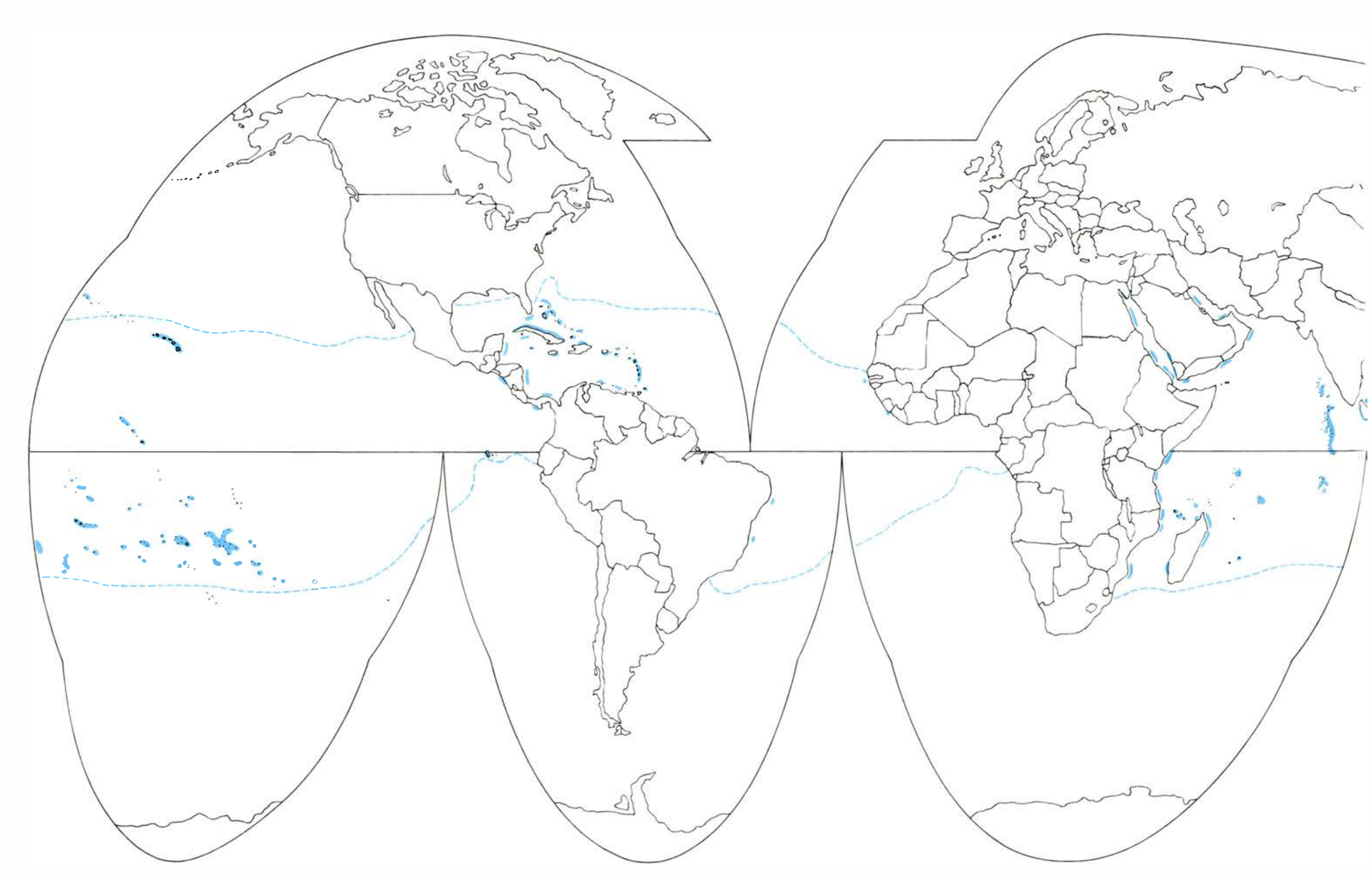
Por razones que no vienen al caso, la fotosíntesis toma el carbono 12 de manera ligeramente más rápida que lo hace con el carbono 13. De aquí que la materia orgánica sintetizada por las zooxantelas tendrá una relativa preponderancia de carbono 12, y se irá produciendo asimismo un acúmulo de compuestos de carbono enriquecidos en carbono 13. Precisamente, a partir de los compuestos de este acúmulo se forma el esqueleto de carbonato cálcico del coral. Determinando las cantidades relativas de los dos isótopos con un espectrómetro de masas se estimó que alrededor de los dos tercios del carbón que se absorbe en la fotosíntesis y en la calcificación se recicla a partir del dióxido de carbono procedente de la respiración del pólipo coralino, tomándose el tercio restante del agua de mar.

La materia orgánica que las zooxantelas transmiten es sólo una de las tres fuentes principales de la alimentación de las madréporas. Los corales son carnívoros eficaces, que inmovilizan a los animales del plancton con las células urticantes de sus tentáculos o los atrapan sobre filamentos de mucus que después son reingeridos. Un pólipo puede detectar químicamente una partícula alimenticia en potencia, y responde extendiendo sus tentáculos, abriendo su boca o evaginando sus filamentos mesentéricos. James Porter, de la Universidad de Georgia, ha analizado el contenido de la cavidad gástrica de los corales y ha encontrado que los pólipos se alimentan principalmente de diminutos crustáceos y animales vermiformes del plancton, que se ocultan en los intersticios del arrecife durante el día y salen al amanecer y al atardecer.

Los estudios con compuestos marcados radiactivamente han demostrado asimismo que los corales pueden absorber materia orgánica disuelta a través de la pared del cuerpo. Puesto que los corales se alimentan activamente de plancton, toman nutrientes del agua de mar y absorben productos químicos liberados por sus zooxantelas, desempeñan a la vez varios papeles ecológicos: productor primario, consumidor primario, detritívoro y carnívoro. Esta compleja red alimentaria reduce su dependencia de cualquier fuente alimentaria única, que podría hallarse sujeta a variaciones al azar al cambiar las condiciones ambientales.

Calcificación en los corales

En los corales el crecimiento se realiza mediante un aumento de la masa del esqueleto calcáreo y el tejido vivo que lo recubre. El esqueleto de los corales está compuesto, en su integridad, por aragonito, una forma cristalina fibrosa del carbonato cálcico ($CaCO_3$); la calcita, la forma cristalina más común del carbonato cálcico, no está presente. En el arrecife hay muchas algas que depositan asimismo aragonito o una forma más soluble de calcita con un elevado contenido en magnesio. Trabajando en Bermuda, Heinz A. Lowenstam, del Ins-



LOS ARRECIFES CORALINOS del mundo (*color*) son de tres tipos básicos: atolones, arrecifes barrera y arrecifes costeros. Los arrecifes de las Antillas son, sobre todo, costeros. Los corales arrecifales se encuentran únicamente en las aguas tropicales soleadas (*líneas de trazos*) porque su

tituto de Tecnología de California, demostró que algunos organismos calcáreos tienden a depositar calcita, menos soluble, en las estaciones frías, y aragonito, más soluble, en las cálidas, pero se desconocen todavía los mecanismos a través de los cuales regulan la mineralogía de su esqueleto.

Los pólipos coralinos absorben iones calcio del agua de mar y los transfieren mediante difusión y gracias a un mecanismo activo de bombeo al lugar de calcificación. Los iones calcio constituyen un regulador bioquímico de primer orden del metabolismo celular, y deben mantenerse a niveles extremadamente bajos para que las células de un tejido funcionen. Aunque los tejidos de las madréporas tienen una concentración total de calcio similar a la del agua de mar, la concentración de los iones libres es mucho más reducida porque la mayor parte del calcio se halla ligado a membranas o a moléculas orgánicas. Lothar Böhm, trabajando en nuestro laboratorio de la Universidad de las Antillas en Jamaica, ha demostrado que el calcio ligado a estos complejos orgánicos tiene una elevada tasa de renovación.

Uno de nosotros (Nora I. Goreau), trabajando en colaboración con Raymond Hayes, de la Facultad de Medicina del Morehouse College, en Atlanta, hizo recientemente detallados estudios de pólipos coralinos al microscopio electrónico. En el curso de estos estudios se observaron, en la capa celular externa del pólipo, minúsculos cristales de carbonato cálcico incluidos en el interior de vesículas rodeadas por membrana. Los cristales son expulsados a través de la membrana hasta el esqueleto del coral, donde actúan como núcleos de un crecimiento cristalino continuo. Este trabajo puede servir para clarificar los mecanismos básicos de calcificación en las células de gran variedad de organismos, en particular debido a que los corales carecen de los controles hormonales sobre la calcificación que complican estos mecanismos en organismos más avanzados.

capacidad de depositar sus esqueletos calizos depende de su simbiosis con las zooxantelas.

El mayor obstáculo para el estudio de la fisiología de la calcificación en las madréporas ha sido la dificultad de mantener corales vivos y saludables en acuarios de laboratorio durante el tiempo suficiente para realizar medidas precisas de la absorción de calcio. Uno de nosotros (Thomas F. Goreau) salvó el problema midiendo la calcificación *in situ*, en el arrecife coralino vivo. Esto se realizó proporcionando al coral calcio en forma del isótopo radiactivo calcio 45, y midiendo la absorción del calcio radiactivo en el esqueleto de la madrépora. El método es tan sensible que puede detectarse crecimiento en corales que han estado expuestos al calcio radiactivo sólo durante unas cuantas horas, que es lo que hace posible los estudios de campo.

Estos estudios han demostrado que aunque los corales constructores de arrecifes crecen en condiciones bastante uniformes de temperatura, iluminación y circulación de agua, hay diferencias muy grandes en las tasas de crecimiento de las distintas especies. Las tasas más elevadas se encuentran invariablemente en los corales ramificados, como en los corales de cuerno de alce y de cuerno de ciervo (*Acropora*), antillanos. *Millepora*, el coral de fuego, ocupa un segundo lugar muy próximo, y los Porítidos (corales digitiformes) el tercero. Los corales macizos crecen más lentamente. En los corales ramificados la mayor parte del crecimiento tiene lugar en las puntas de las ramas, y en casi cualquier lugar de las partes más viejas de la colonia se desarrollan nuevas ramas.

Factores que influyen sobre la calcificación

Un factor crucial que influye sobre la tasa de calcificación es la conversión de dióxido de carbono (CO_2) respiratorio en ácido carbónico (H_2CO_3), que a su vez se convierte en iones bicarbonato (HCO_3^-) y carbonato ($CO_3^{=}$). El enzima responsable de la adición de agua al dióxido de carbono para formar ácido carbónico es la anhidrasa carbónica, que se encuentra en concentraciones elevadas en los corales. La formación subsiguiente de iones bicarbonato y carbonato es rápida y no requiere ser catalizada por ningún enzima. Las drogas que inhiben la anhidrasa carbónica producen un espectacular descenso en la tasa de calcificación.

Por término medio, el crecimiento del esqueleto coralino es 14 veces más rápido a la luz que en la oscuridad, y puede reducirse mediante la acción de drogas que bloquean la fotosíntesis. Las propias variaciones diarias en la intensidad de la luz tienen un efecto mensurable sobre la tasa de calcificación: la absorción de calcio es más rápida al mediodía de un día claro y soleado, se reduce en un 50 por ciento en un día nublado y en cerca del 90 por ciento en oscuridad total. La intensidad de la luz ambiente disminuye asimismo con la profundidad: el flujo luminoso a una profundidad de 60 metros es sólo el 4 por ciento del flujo en superficie. Como resultado, la tasa a la que se deposita calcio en el esqueleto del coral es probable que decrezca rápidamente con el aumento de profundidad.

El hecho sorprendente de que el crecimiento del coral dependa de la intensidad de la luz ambiente se observa sólo cuando hay zooxantelas presentes. Si se extraen las algas simbiontes (manteniendo la colonia de coral en la oscuridad durante varios meses), la tasa de calcificación es baja y ya no se ve afectada por cambios en la intensidad de la luz, que es la regla general en los corales no simbióticos. ¿De qué manera aumentan las zooxantelas la tasa de calcificación? La respuesta parece ser que la fijación de dióxido de carbono por parte de las algas produce un aumento en la concentración de iones carbonato en las células del pólipo coralino a través de una serie de reacciones químicas relacionadas, que hacen aumentar el *p*H del líquido celular, de modo que éste es más alcalino. Al precipitar sus iones carbonato en exceso en forma de carbonato cálcico insoluble, el pólipo puede restablecer su *p*H al nivel normal y, al mismo tiempo, construir su esqueleto calizo. Las zooxantelas pueden asimismo estimular la calcificación indirectamente, al aumentar la cantidad de energía libre disponible para el transporte activo de iones calcio hasta el lugar de la calcificación. Por tanto, las algas actúan de manera sinérgica con la anhidrasa carbónica

VISION SUPERIOR
DE LA COPA ESQUELETICA

TENTACULOS

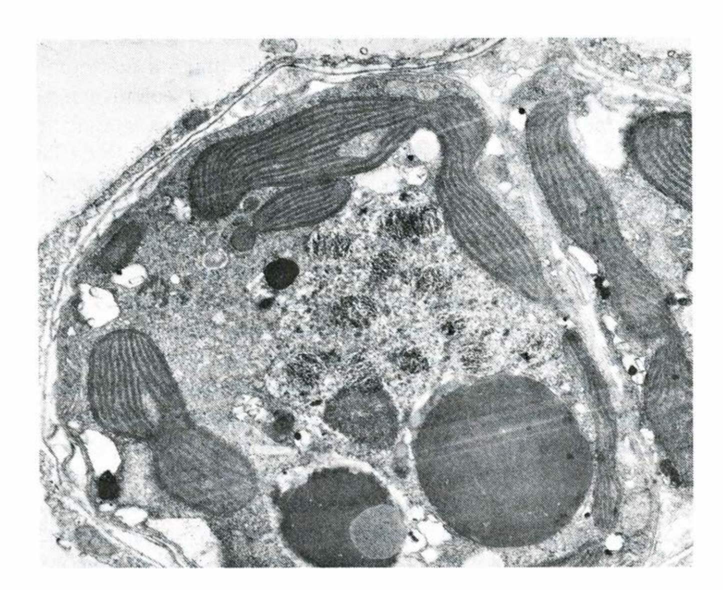
FARINGE

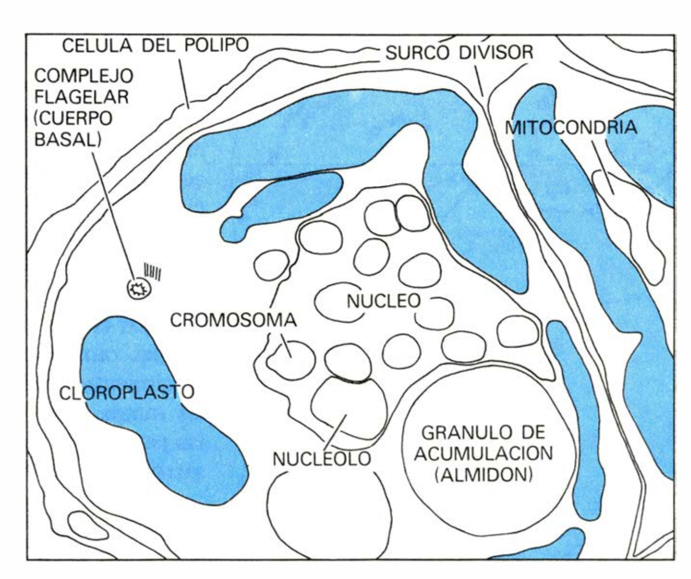
SEPTO

FILAMENTOS
MESENTERICOS

ZOOXANTELAS

ANATOMIA DEL POLIPO CORALINO. Es simple: el animal consiste básicamente en un saco contráctil formado por tres capas tisulares. El cuerpo cilíndrico está coronado por una boca central rodeada por tentáculos; desde la boca, una faringe muscular desciende hacia la cavidad digestiva central, que se halla conectada al cuerpo por una serie de particiones verticales. Los bordes libres de estas particiones se extienden en forma de filamentos mesentéricos. Las células del revestimiento de la cavidad digestiva albergan las algas simbiontes, que viven, fotosintetizan y se dividen en el interior de las células del patrón. Los pólipos se asientan en copas protectoras de caliza que constan de una disposición radial de placas verticales, que se interdigitan con las particiones del pólipo. Cada pólipo coralino deposita nuevos pisos bajo él, al tiempo que va creciendo hacia arriba.



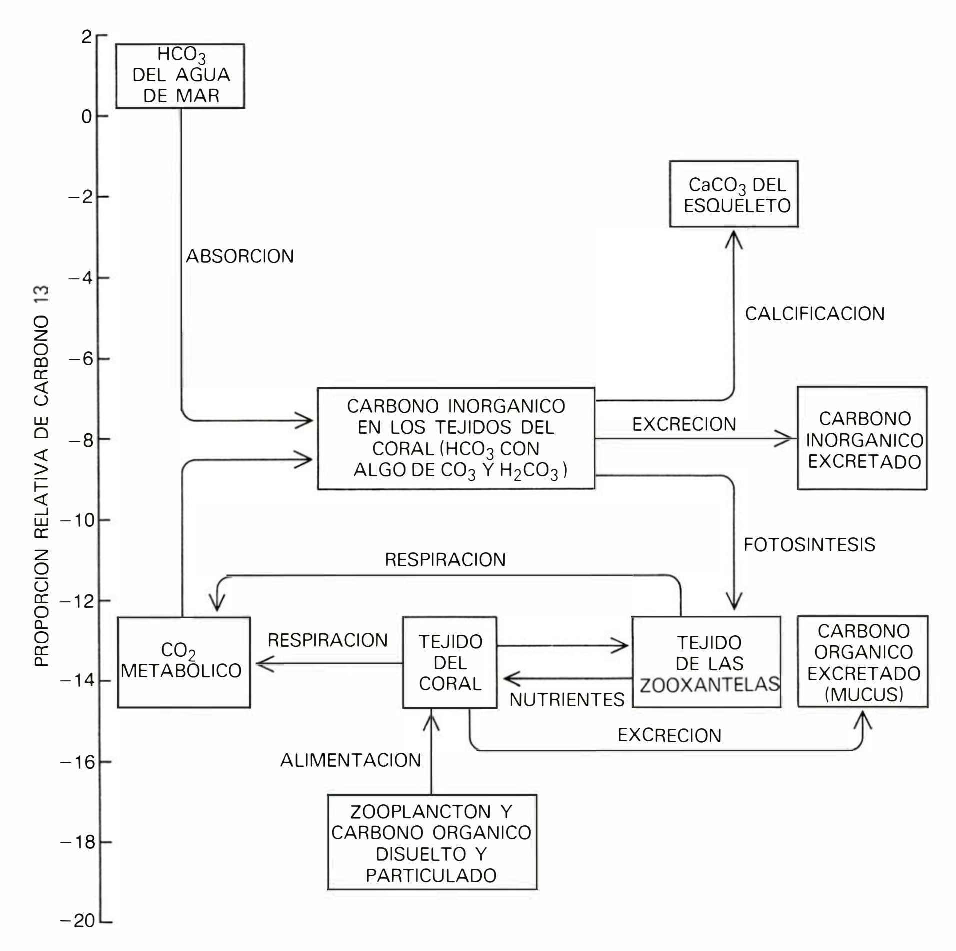


CELULA ALGAL (ZOOXANTELA) dividiéndose, aumentada 25.000 diámetros en la micrografía electrónica de la izquierda. Los sacos estriados dentro de las células son secciones de un cloroplasto grande único, en el que tiene lugar la fotosíntesis; los demás orgánulos celulares se indican en el mapa de la derecha. Las zooxantelas aumentan mucho la eficiencia metabólica del patrón coralino al absorber los productos de desecho de la respiración del coral y al reciclar algunos de ellos en nueva materia orgánica. También "pasan" nutrientes esenciales a los pólipos coralinos y acrecientan la tasa de calcificación. La micrografía electrónica de la izquierda se debe a Robert K. Trench, de la Universidad de California en Santa Barbara.

para aumentar la formación de carbonato cálcico. La calcificación puede tener lugar en ausencia de fotosíntesis algal, pero únicamente a una tasa reducidísima.

El hecho de que la calcificación en las madréporas esté controlada biológicamente viene indicado además por las variaciones estacionales en la tasa de crecimiento. Estas variaciones se reflejan en las medidas que efectuó uno de nosotros (Thomas J. Goreau) de la concentración de magnesio, un metal traza, y de los isótopos pesados y ligeros de carbono y de oxígeno en las bandas estacionales de crecimiento. Una vez que se comprendan mejor las influencias ambientales y fisiológicas que afectan al crecimiento del coral, las variaciones en la composición de los esqueletos coralinos proporcionarán un registro químico tan detallado de los ambientes pretéritos como el que puede obtenerse de los anillos de crecimiento de los árboles.

El efecto sinérgico de las zooxantelas sobre la tasa de calcificación fue desde luego un factor decisivo en la evolución de los arrecifes coralinos. El desarrollo de inmensas comunidades coralinas a pesar de la acción demoledora del mar embravecido sólo resultó posible cuando los procesos de deposición de carbonato cálcico se hicieron suficientemente eficaces para que la tasa de deposición superara la tasa de pérdida debida al desgaste físico y biológico.



CICLO DEL CARBONO, esquematizado, entre las zooxantelas, la madrépora patrón y el ambiente. Sobre un eje vertical se indican las distintas reservas de carbono según la relación de los dos isótopos estables de este elemento: carbono 12 y carbono 13. La posición de cada reserva es, pues, una indicación de la importancia relativa de los procesos mediante los cuales cada una de ellas gana o pierde carbono y de la intensidad con que cada uno de estos procesos utiliza preferentemente uno de los dos isótopos. Por ejemplo, debido a que la fotosíntesis absorbe más rápidamente el carbono 12 que el carbono 13, deja tras sí una reserva enriquecida en carbono 13, de la que se forma el carbonato cálcico del esqueleto madrepórico. Alrededor de los dos tercios del carbono utilizado en la fotosíntesis y en la calcificación se reciclan a partir del dióxido de carbono (CO_2) de la respiración. El nivel del isótopo estable carbono 13 en el tejido del coral refleja cuál sea la composición de sus fuentes nutritivas.

Arquitectura del arrecife

Quizá los pólipos coralinos no lleguen a dominar la biomasa (la masa total de materia viva), la productividad biológica o incluso la calcificación en todas las partes de un arrecife de coral. No obstante, la existencia de muchas comunidades animales y vegetales del arrecife se basa en la capacidad de la madrépora para construir una estructura robusta y resistente al oleaje. Las interacciones dinámicas de los procesos geológicos y biológicos que controlan el crecimiento de los arrecifes de coral están bien ilustradas en el arrecife costero que se extiende a lo largo de 240 kilómetros por la costa septentrional de Jamaica, y que hemos venido estudiando durante los últimos 28 años.

La principal característica estructural del arrecife vivo es un bastión de coral que llega casi hasta la superficie del agua. Está constituido por cabezas de corales redondeados y macizos y corales ramificados robustos, que forman una empalizada rígida y cavernosa de esqueletos de coral anastomosados. Sobre este entramado viven madréporas más pequeñas y frágiles y grandes cantidades de algas calcáreas verdes y rojas. La biomasa de estas algas es pequeña comparada con la de los corales, pero su productividad y su tasa de renovación son tan altas que la arena formada por sus restos esqueléticos constituye la mayor parte del carbonato cálcico depositado en el arrecife.

Centenares de especies de organismos incrustantes viven sobre la parte superior del entramado coralino, uniendo con sus finas extensiones las ramas de coral. Innumerables peces e invertebrados se ocultan asimismo en los escondrijos y grietas del arrecife, y algunos de ellos sólo salen por la noche. Además, los organismos sésiles cubren virtualmente todo el espacio disponible sobre la cara inferior de las placas de coral y sobre los esqueletos de las madréporas muertas.

La cresta arrecifal corre paralela a la costa, en algunos lugares tocando la orilla y en otros encerrando una laguna arenosa de unos cinco metros de profundidad y de hasta unos cuantos cientos de metros de anchura. Esta zona se halla protegida de la resaca y está salpicada de bloques de coral aislados. La laguna está dominada por manchas de algas calcáreas y una comunidad de animales bentónicos, en especial erizos de mar y holoturias, que se sustentan filtrando la materia orgánica contenida en los sedimentos o en la capa de agua situada inmediatamente por encima de ellos. Muchos de estos organismos ramonean algas filamentosas; si de una zona de la laguna se extraen los animales herbívoros, después de sólo unos cuantos días se forma una densa alfombra de algas. Las actividades de construcción de madrigueras y de agitación que producen los herbívoros son importantes porque liberan los nutrientes creados por la descomposición bacteriana de la materia orgánica enterrada en los sedimentos. Densas "praderas" de la fanerógama marina *Thalassia* forman hábitats especiales que albergan a su

propia comunidad de erizos de mar, caracolas y muchas otras especies.

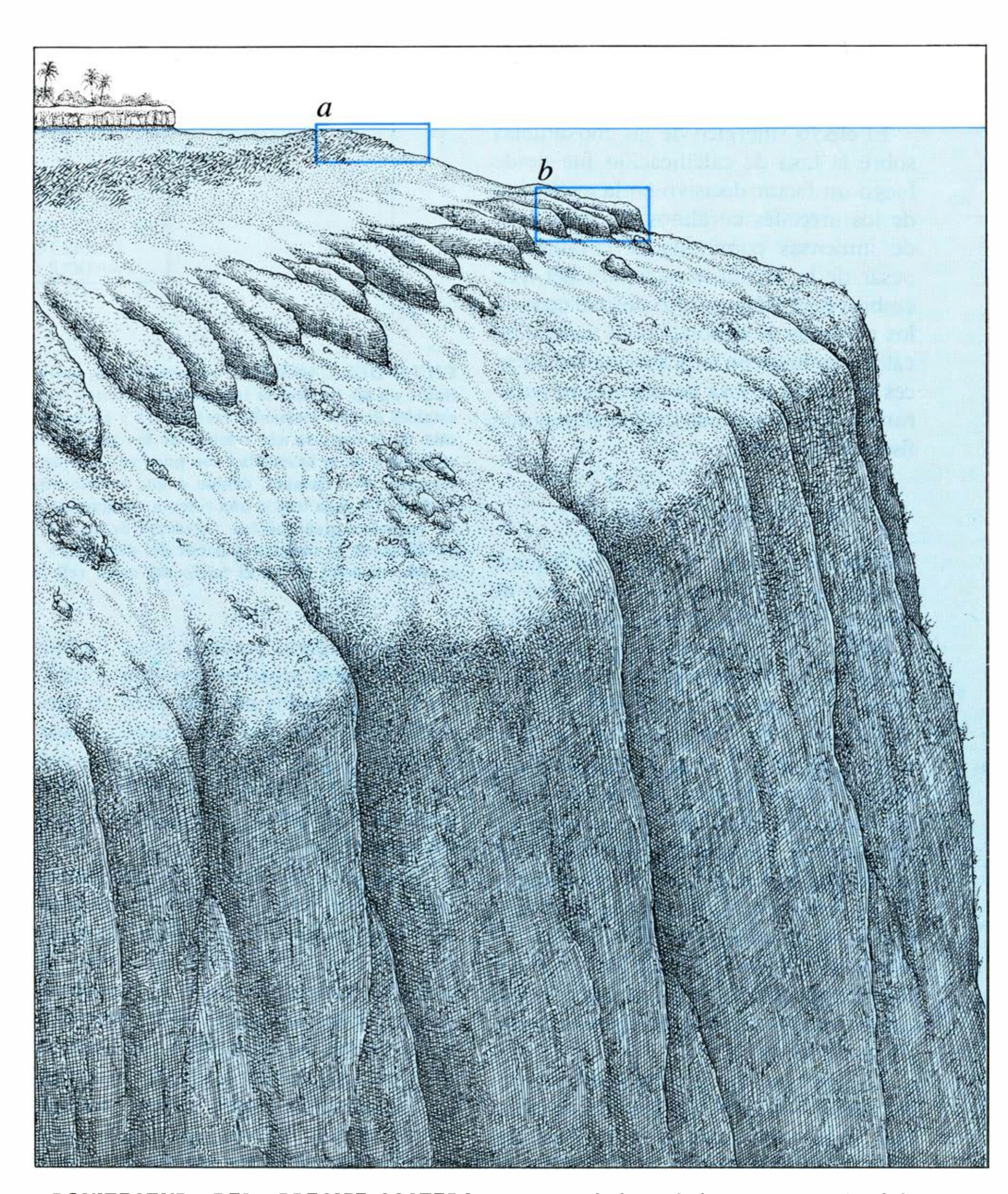
Del lado del mar de la cresta arrecifal se encuentra el antearrecife, en el que los corales tapizan casi todo el fondo marino. Los corales forman grandes contrafuertes separados por estrechos canales arenosos, a cuyo través pasa un flujo continuo de sedimento fino que se origina por la desintegración de las madréporas, de las algas calcáreas y de otros organismos muertos. Los canales parecen estrechos cañones sinuosos con paredes verticales de matas de sólido coral. Pueden tener hasta 10 metros de profundidad, y algunos de ellos están completamente techados con coral. Esta espectacular interdigitación de contrafuertes y canales disipa la energía del oleaje y al mismo tiempo permite el libre flujo de sedimentos que, de otra manera, ahogarían el crecimiento del coral.

Crecimiento del arrecife

Debajo de la zona de contrafuertes se encuentra una terraza de coral, un talud de arena con pináculos coralinos aislados, después otra terraza y, finalmente, una pared casi vertical que cae hacia la oscuridad de las grandes profundidades. La distribución de las especies de madréporas y de las otras comunidades animales en el arrecife sigue una zonación debida a la profundidad, característica que permite a los paleontólogos que estudian la sección de un arrecife antiguo que ahora se halla en tierra firme estimar de forma precisa la profundidad original de esta sección a partir de los animales fósiles a ella asociados. En aguas de profundidad superior a los 100 metros pocas algas o corales simbiontes crecen bien debido a los reducidos niveles de luz, y la fauna está dominada por animales que capturan o filtran los detritos orgánicos que caen desde el arrecife situado más arriba. Los animales detritívoros incluyen las verdaderas esponjas, los antipatarios (los "corales preciosos") y las gorgonias (los abanicos de mar). También son comunes aquí las escleroesponjas, un grupo antiguo que en el pasado geológico fueron importantes constructores de arrecifes pero que durante mucho tiempo se pensó que se habían extinguido hace cientos de millones de años. Nuestros estudios en inmersión en los arrecifes profundos de Jamaica han demostrado que estas esponjas se encuentran vivas y bien, pero desplazadas a hábitats más profundos por los corales, de crecimiento más rápido, que evolucionaron más tarde.

El crecimiento del arrecife es el resultado de una relación dinámica entre la extensión hacia arriba del entramado de coral y el acarreo hacia afuera de un volumen mucho mayor de detritos de grano fino. La salida de sedimentos del arrecife se consigue sobre todo mediante el flujo y el arrastre gravitacionales, ya sea hacia el interior de la laguna o a lo largo de los canales de la zona de contrafuertes hacia aguas profundas. Puede ocurrir asimismo que montones inestables de coral crezcan hasta que se desmoronan bajo su propio peso y se derrumban. Cuando se exploró la región inferior del arrecife jamaicano en el submarino de investigación *Nekton Gamma II* a profundidades de más de 200 metros, en la base del escarpado talud se observaron enormes acúmulos de sedimento y gigantescos bloques de sólido arrecife; pudieron haberse desmoronado por causa de los terremotos. Estas dislocaciones crean sustratos nuevos para los organismos incrustantes y colaboran al establecimiento de comunidades coralinas en los escarpados taludes inferiores, en especial las colonias laminares y espiraladas de *Agaricia*.

Otros dos procesos importantes influyen sobre el crecimiento del arrecife: la erosión biológica y la litificación submarina. Muchas especies de algas filamentosas, hongos, esponjas, gusanos marinos, crustáceos y moluscos perforan los esqueletos de los corales, excavando agujeros mediante raspado mecánico o disolución química. La más común es la esponja perforante *Cliona*, que extrae minúsculas briznas de carbonato cál-



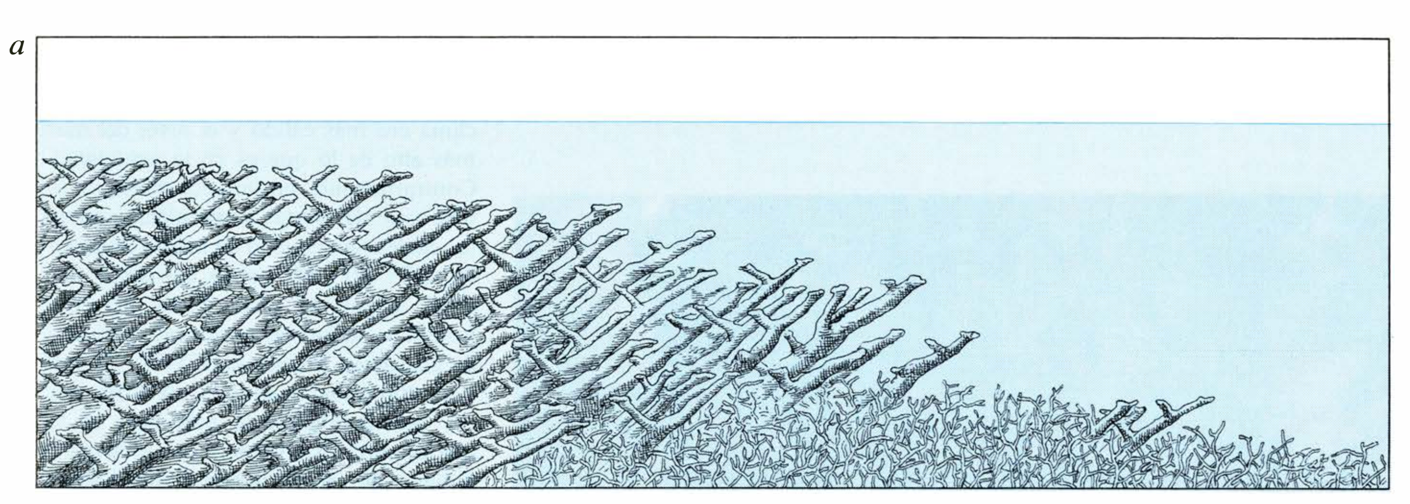
ARQUITECTURA DEL ARRECIFE COSTERO que corre a lo largo de la costa septentrional de Jamaica, ilustrada en esquemas tridimensionales. Pueden distinguirse varias zonas a partir de la estructura del arrecife, de su profundidad y de las comunidades animales y vegetales asociadas. La cresta o corona arrecifal se extiende hasta una profundidad de unos 15 metros y comprende el muro coralino, somero, y la zona de rompiente (*a*). El antearrecife se extiende de los 15 a los 30 metros. Esta región es un ambiente de energía media, con una intensidad de luz ambiente del orden del 25 por cien de la que hay en superficie. La zona de contrafuertes (*b*), en la que los contrafuertes de coral alternan

cico; estos restos son el principal componente de los sedimentos finos. *Cliona* puede acribillar de agujeros el esqueleto de un coral sin dañar a los pólipos coralinos vivos. En las aguas más profundas muchos corales crecen formando láminas planas y delgadas para maximar la superficie de captación de luz, y por ello son bastante susceptibles a la erosión por parte de los perforadores, que pueden hacer que las madréporas se rompan y caigan talud abajo. Sin embargo, en algunos lugares el coral está tan recubierto por organismos incrustantes que permanece en su lugar aunque ya no se halle directamente unido al arrecife.

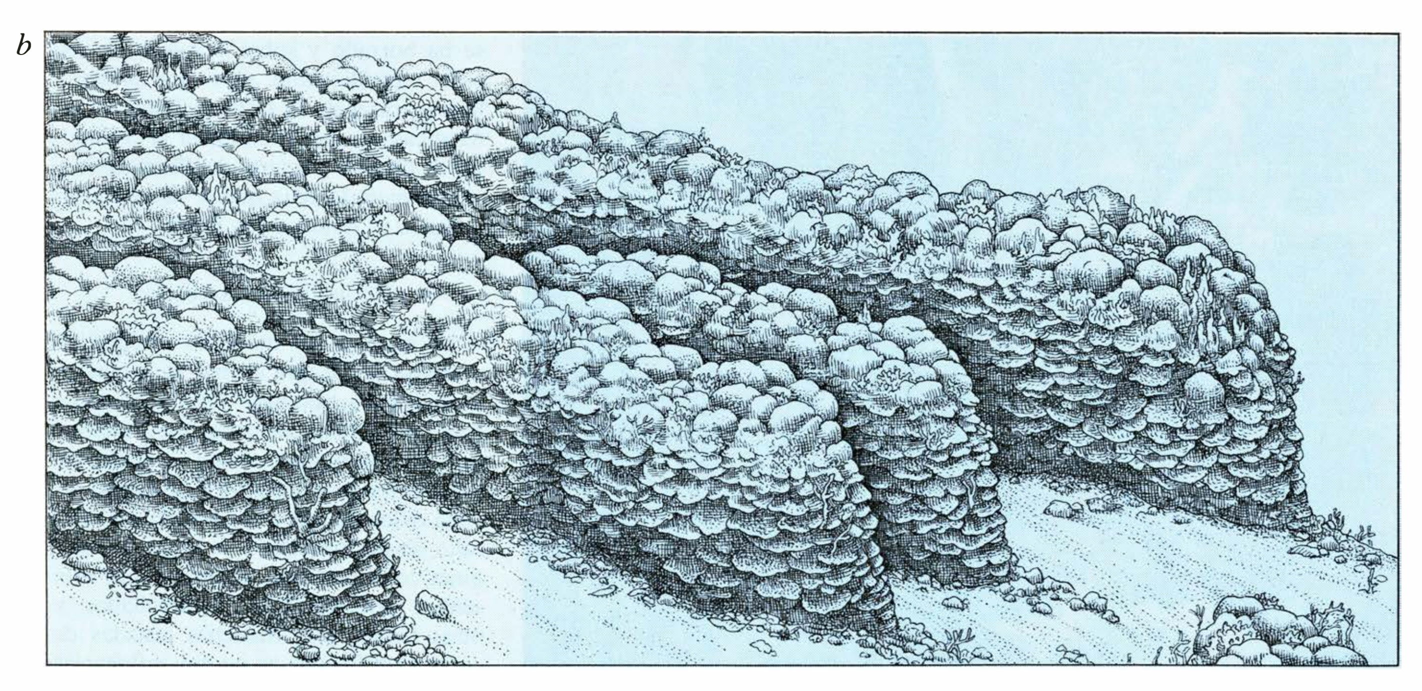
La litificación submarina contrarresta los efectos de la erosión biológica; consiste en la deposición de un cemento carbonatado de grano fino en los poros y cavidades del esqueleto coralino. Los sedimentos que quedan atrapados en el entramado arrecifal son rápidamente unidos por la acción de los organismos incrustantes y del cemento calcáreo. El origen de este cemento todavía no se conoce bien; puede tratarse de un precipitado inorgánico producido por bacterias que viven en las anfractuosidades del arrecife. Estudios realizados en el Discovery Bay Marine Laboratory, en Jamaica, conjuntamente con Lynton S. Land, de la Universidad de Texas en Austin, demostraron que una vez el cemento se ha endurecido es perforado y vuelto a llenar; los agujeros rellenados son aparentes cuando se examinan secciones delgadas del agregado bajo el microscopio. La litificación submarina produce la acreción del antearrecife por su parte externa, y estabiliza el perfil abrupto de la escarpada pared de bajada. El crecimiento de los arrecifes es, pues, el producto de un equilibrio dinámico entre el crecimiento del entramado, el transporte de sedimentos, la bioerosión producida por los perforadores, la destrucción mecánica y la litificación submarina; la importancia relativa de estos factores varía de un arrecife a otro.

El arrecife vivo es, básicamente, un revestimiento que crece unos pocos centímetros por año, situado encima de una compleja topografía de arrecifes ancestrales superpuestos. En Jamaica se han desarrollado por lo menos nueve metros de arrecife desde que el actual nivel del

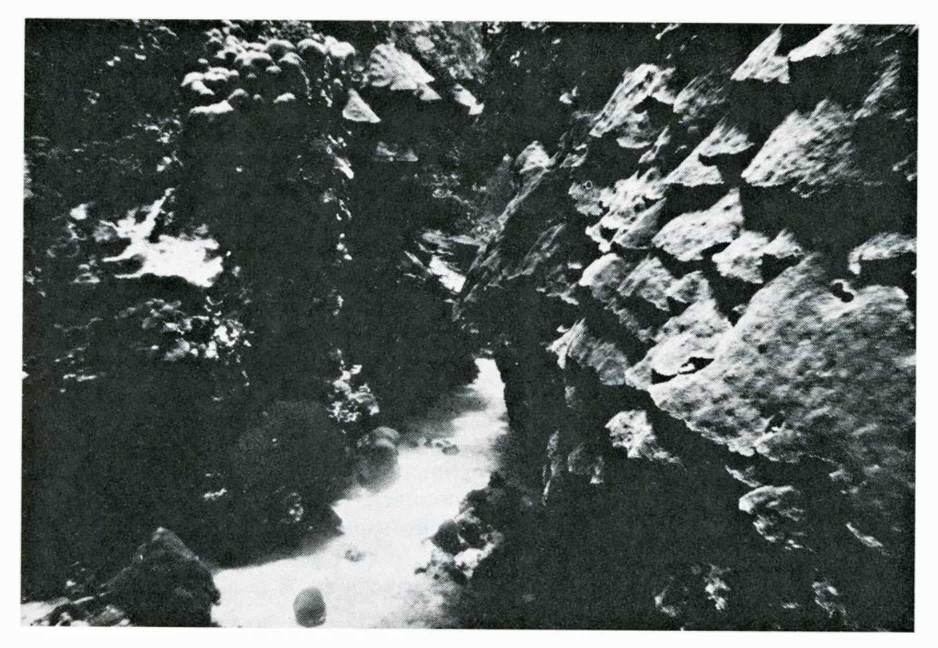
a



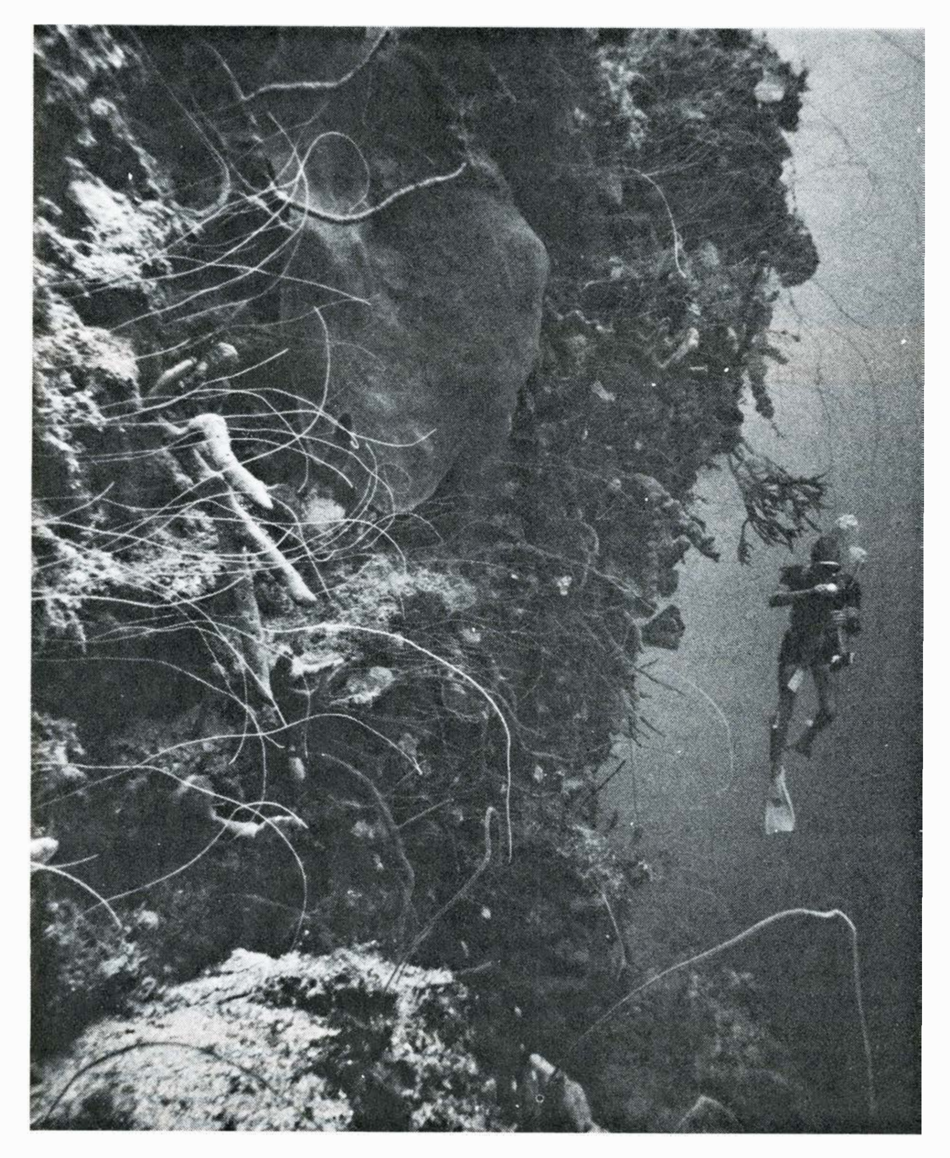
b



con cañones de arena, sirve para disipar la energía mecánica de las olas y permite el flujo, arrecife abajo, de finos sedimentos que, de otro modo, impedirían el crecimiento de las madréporas. Las colonias coralinas son todavía variadas pero de tamaño menor, y gran parte del espacio disponible está ocupado por algas calcáreas productoras de arena, esponjas y grandes gorgonias (abanicos de mar). El antearrecife profundo se extiende desde los 30 a los 70 metros. Esta zona tiene una topografía muy inclinada y se halla poco iluminada, con un flujo luminoso que es de alrededor del 5 por ciento del que llega a la superficie. A profundidades inferiores a los 30 metros, las matas de coral se hallan más dispersas, con una reducción progresiva del número de especies y del tamaño y densidad de las colonias. Hay también un extenso transporte de sedimentos desde las zonas más someras. Más allá del antearrecife profundo la pared vertical cae hacia la oscuridad. En los arrecifes profundos existen escleroesponjas que se suponían extinguidas.



CAÑON DE CONTRAFUERTES entre dos paredes de coral, fotografiado a una profundiad de unos 12 metros en aguas de la costa de Jamaica. La pared de la derecha está cubierta por colonias del coral *Montastrea annularis*. La forma de las colonias ayuda a maximar su superficie captadora de luz.



PARED ESCARPADA del arrecife costero de Jamaica, fotografiada a una profundidad de unos 40 metros. El talud antearrecifal está cubierto por acúmulos de esponjas, gorgonias y corales látigo.

mar se estabilizó hace unos 5000 años. Los arrecifes antiguos permanecen, y proporcionan un registro de los cambios en el nivel del mar y del levantamiento continental debido a los movimientos de las placas tectónicas.

La elevación y el descenso del nivel del mar a lo largo de los últimos millones de años han sido producidos por los cambios en el volumen del agua retenida en los glaciares y campos de hielo continentales durante las eras glaciales del Pleistoceno. Cuando en el hemisferio Norte se formaron los campos de hielo, bajó el nivel del mar y los arrecifes de coral quedaron en seco por encima de dicho nivel. En la actualidad, crestas fósiles y desfiladeros cortados por las olas indican el antiguo nivel del mar. En Jamaica, Barbados, Nueva Guinea y en otras costas coralinas se encuentra una sucesión de arrecifes varados; estos arrecifes se formaron hace 80.000, 105.000, 125.000 y 200.000 años, cuando el clima era más cálido y el nivel del mar más alto de lo que es en la actualidad. Contrariamente, en Jamaica puede verse una serie de crestas sumergidas y sobre las cuales han crecido de nuevo los corales, a 25, 40 y 60 metros por debajo del nivel actual del mar. Estos arrecifes hundidos se formaron durante periodos de glaciación intensa hace 8000, 11.000 y 14.000 años, cuando el nivel del mar era considerablemente más bajo de lo que es ahora. De manera que el arrecife antiguo es un palimpsesto difícilmente visible situado por debajo del arrecife vivo, como un manuscrito medieval que se ha borrado y sobre el que se ha escrito repetidamente, pero que presenta ligeras trazas de su historia. Estas características ayudan a establecer la cronología de las eras glaciares del Pleistoceno y el volumen de agua que se añadió a los océanos al fundirse los hielos.

Ecología del arrecife

La historia del moderno arrecife jamaicano desde que el mar se estabilizó en su nivel actual hace 5000 años no ha sido suficientemente larga para establecer una comunidad clímax, es decir, un ecosistema en equilibrio. Este hecho resulta evidente por el desarrollo casi fortuito de los arrecifes a lo largo de cualquier costa coralina: algunas zonas presentan arrecifes bien desarrollados y otras tienen sólo manchas aisladas de coral. Con frecuencia no hay influencias ambientales o factores catastróficos (como terremotos u olas de marea) evidentes que expliquen estas diferencias de desarrollo. Más bien parece que las variaciones casuales en el establecimiento de las larvas nadadoras de los corales y en

el crecimiento desempeñan el principal papel a la hora de determinar la formación de los arrecifes, y que simplemente no ha habido tiempo suficiente para que los corales ocupen todos los hábitats favorables.

El papel del azar en la instalación de los corales se refleja asimismo en la variabilidad de las principales especies que desempeñan los mismos papeles estructurales en un arrecife determinado. En algunos arrecifes de Jamaica el coral dominante es el coral ramificado *Montastrea annularis*, pero en hábitats similares el mismo nicho lo ocupa una especie distinta, *Agaricia tenuifolia*, que construye colonias de forma, tamaño y orientación idénticas. De ahí que en la creación de diversidad en un arrecife coralino la variación histórica sea tan significativa en muchos arrecifes como lo es el acercamiento a un equilibrio en el que coexisten muchos organismos especializados.

La multiplicidad de hábitats localizados y de especies en el arrecife dan a la comunidad arrecifal un cúmulo de interacciones intra- e interespecíficas cuya complejidad puede comprenderse sólo de manera muy tenue. Unicamente después de años de experiencia de campo puede obtenerse una comprensión intuitiva de las principales interacciones. Incluso entonces, uno puede centrarse en tan pocos componentes de la comunidad que es fácil pasar por alto papeles significativos que desempeñan organismos raros, no examinados o desconocidos.

La competencia intensa por el alimento y por el espacio en el hábitat arrecifal ha dado origen a una gran variedad de estrategias de supervivencia. Por ejemplo, Jeremy Jackson y sus discípulos de la Universidad Johns Hopkins han demostrado que muchos organismos incrustantes poseen toxinas específicas con fines defensivos u ofensivos. Los corales que crecen juntos compiten por el espacio, y algunas especies pueden extraer los filamentos mesentéricos del tubo digestivo para matar a los pólipos de las colonias adyacentes. Judith Lang, trabajando en Discovery Bay, ha demostrado que entre las especies de madréporas hay una jerarquia de agresión, de modo que los corales de crecimiento lento y agresivos pueden evitar quedar recubiertos por los corales de crecimiento más rápido aunque menos agresivos. Este proceso puede conducir a un aumento de la diversidad específica. En algunos casos, sin embargo, resulta precisamente todo lo contrario: James Porter ha observado que en los arrecifes de la costa pacífica de Panamá el coral abrumadoramente dominante, *Pocillopora damicornis*, constituye, a la vez, la especie de crecimiento más rápido y la más agresiva.

El ramoneo que sobre los tejidos de las algas y de los corales efectúan los peces, los erizos de mar y otros animales tiene dos efectos importantes. El pastoreo selectivo puede evitar que unas cuantas especies dominantes de algas se multipliquen y excluyan a las especies más marginales, de modo que puede existir una cierta diversidad de especies. Los experimentos en los que se excluyen a los organismos ramoneadores de una zona del arrecife concluyen, por lo general, con unas pocas especies vegetales dominantes, con densidades elevadas, que excluyen a las demás especies y que son raras en circunstancias ordinarias. Los ramoneadores que arrancan tejidos de los sustratos duros crean asimismo nuevas superficies sobre las que pueden crecer nuevas algas e instalarse las larvas de organismos sésiles. Leslie S. Kaufman, de la Johns Hopkins, ha descubierto que algunas especies de peces destruyen de manera sistemática fragmentos de tejido de coral, de modo que sobre el esqueleto coralino desnudo puedan crecer "granjas" de algas. Los peces, que pastan las algas, expulsan a los intrusos de su territorio, peces mucho mayores y buceadores humanos incluidos. No se sabe el daño que este "aclarado" biológico, comparado con el que producen los hundimientos y las tormentas, causa en el arrecife.

Queda mucho por saber acerca de los ciclos de los nutrientes y de la energía en los arrecifes. La riqueza de los procesos biológicos del arrecife frente a la pobreza de los nutrientes disueltos en las aguas tropicales superficiales es una prueba de que existe un eficiente ciclo interno de los nutrientes en el ecosistema arrecifal, pero todavía ha de investigarse el asunto con el debido detalle. Se cree que el principal nutriente limitante en los océanos es el nitrógeno, y en los arrecifes de coral las algas cianofíceas filamentosas fijan en formas utilizables grandes cantidades del nitrógeno atmosférico disuelto en el agua de mar. Otra fuente de nitratos es la oxidación de amonio que realizan las bacterias en el curso de la descomposición de la materia orgánica en los sedimentos de la laguna arrecifal. Trabajos recientes indican que la oxidación de amonio a nitrato es particularmente intensa en los sedimentos orgánicos de grano fino retenidos por los rizomas de las praderas de fanerógamas marinas.

Los arrecifes coralinos del Atlántico, del Caribe y del Indopacífico no difieren fundamentalmente en sus formas estructurales, sus hábitats y las interacciones de sus especies, aunque los organismos

que ocupan nichos ecológicos específicos varían mucho de un océano a otro e incluso entre un arrecife y otro. Entre el Pacífico y el Caribe, sin embargo, hay una diferencia fundamental: en el Pacífico el crecimiento activo de los corales se realiza únicamente hasta los 60 metros de profundidad, y en el Caribe hasta los 100 metros. La reducida extensión en profundidad de los corales del Pacífico puede deberse en parte a infestaciones periódicas de la estrella de mar de corona de espinas (*Acanthaster planci*), que se alimenta de los corales evaginando su estómago, extendiéndolo sobre el coral y digiriendo los tejidos de éste. Antes del brote reciente y bien conocido de *Acanthaster*, este organismo se hallaba limitado a aguas más profundas y se le veía raramente. Con posterioridad, una explosión demográfica no explicada dio origen a una escasez alimentaria que forzó a las estrellas de mar a ascender hacia aguas más someras, donde sus efectos destructivos se hicieron sentir muy pronto. El límite inferior de crecimiento arrecifal en el Pacífico puede estar pues afectado por un pastoreo periódico debido a las estrellas de mar. Queda todavía mucho por hacer, sin embargo, para probar esta hipótesis; sobre todo, porque muchos arrecifes del Pacífico presentan síntomas de estar sometidos a una erosión mecánica más intensa que los arrecifes caribes.

Estos aspectos ilustran algunas de las dificultades con que los ecólogos se enfrentan al intentar predecir la estabilidad de las poblaciones arrecifales en respuesta a los cambios ambientales, o la sensibilidad de las redes alimentarias arrecifales a las alteraciones en la abundancia de una especie determinada. Puesto que los arrecifes de coral son centros localizados de elevada productividad biológica, y sus peces de hermosos colores una fuente importante de alimento en las zonas tropicales, muchos biólogos marinos ven con alarma la proliferación de promociones turísticas a lo largo de las costas coralinas en muchas regiones del globo. Estas instalaciones se ven casi siempre acompañadas por un vertido creciente de desechos, por sobrepesca, por el daño físico al arrecife producido por la construcción, el dragado, el llenado y colmatado de tierras, y por la destrucción del arrecife a gran escala para proporcionar a los turistas *souvenirs* y curiosidades para el aparador. En muchas zonas (como en Bermuda, las islas Vírgenes y Hawai, de los Estados Unidos) el desarrollo y los vertidos del alcantarillado han producido una eutrofización extensa, es decir, el recubrimiento y posterior muerte del arrecife por densas matas de algas filamentosas, que a su vez provocan el crecimiento de bacterias consumidoras de oxígeno. Los resultados, que están siendo estudiados de manera intensiva por Stephen V. Smith y sus colegas de la Universidad de Hawai, incluyen un aumento de la sensibilidad de los corales a las enfermedades bacterianas, la muerte del coral vivo y la consiguiente erosión del arrecife, y la producción del maloliente sulfuro de hidrógeno.

Romper una barrera

El proyecto de excavar un nuevo canal al nivel del mar a través del istmo de Panamá ha originado nuevas inquietudes acerca de la viabilidad de los arrecifes de coral y de sus recursos físicos y biológicos, intrincadamente entrelazados. La gran amplitud de las mareas en la zona pacífica del istmo y la reducida amplitud de las mismas en la zona caribe, junto al nivel medio del mar más alto en la costa pacífica, produciría el movimiento efectivo a través del canal de especies marinas del Pacífico hacia el Caribe y el Atlántico. Puesto que los arrecifes del Caribe y del Pacífico han estado aislados evolutivamente durante millones de años, una incursión de especies a una escala tan grande en nuevos hábitats podría permitir que algunas de ellas se multiplicaran y se extendieran sin competencia, con consecuencias ecológicas similares a la multiplicación explosiva de los conejos ingleses introducidos en Australia. Así, la estrella de mar de corona de espinas es común en la costa pacífica del istmo, pero no está presente en la costa caribe, y su expansión a través de un canal situado al nivel del mar podría diezmar los corales del Caribe y del Atlántico. Además, en la costa pacífica del istmo abundan las serpientes de mar venenosas, desconocidas hasta ahora en el Atlántico. Peter Glynn e Ira Rubinoff, del Smithsonian Tropical Research Institute, en Panamá, han advertido que el canal al nivel del mar causaría una perturbación mayor en el ambien natural que la originada por cualquier obra de ingeniería anterior.

El canal a nivel del mar propuesto ilustra no sólo las preocupaciones de los biólogos de los arrecifes de coral, sino también su ignorancia, puesto que sigue siendo difícil predecir los efectos perjudiciales de las actividades humanas en un ambiente tan complejo como el ecosistema arrecifal. De todos modos, no parece excesivamente alarmista advertir del riesgo que significa no apreciar en todo lo que vale la estabilidad y la productividad de la comunidad arrecifal.

Supernovas y formación de estrellas

La muerte explosiva de una estrella de gran masa puede desencadenar el nacimiento de otras estrellas. La teoría y observaciones de restos de antiguas supernovas y meteoritos permiten sustentar esta hipótesis

William Herbst y George E. Assousa

Una supernova, catastrófica explosión de una estrella de gran masa al final de su ciclo de vida, lanza un estrato esférico de gases al medio interestelar. La explosión libera simultáneamente una tremenda cantidad de energía: unos 10^{51} joule (siendo un joule el trabajo necesario para elevar un kilogramo a una altura de 10 centímetros). Hace unos 25 años, Ernst J. Öpik, del Observatorio Armagh en Irlanda del Norte, sugirió que la muerte explosiva de una estrella de gran masa podía desencadenar el nacimiento de otras estrellas. Argüía que la onda de choque procedente de una supernova podía comprimir de manera efectiva una nube difusa de gas interestelar y polvo hasta una densidad suficientemente grande para que la gravedad reuniera la materia dispersa formando una estrella o estrellas.

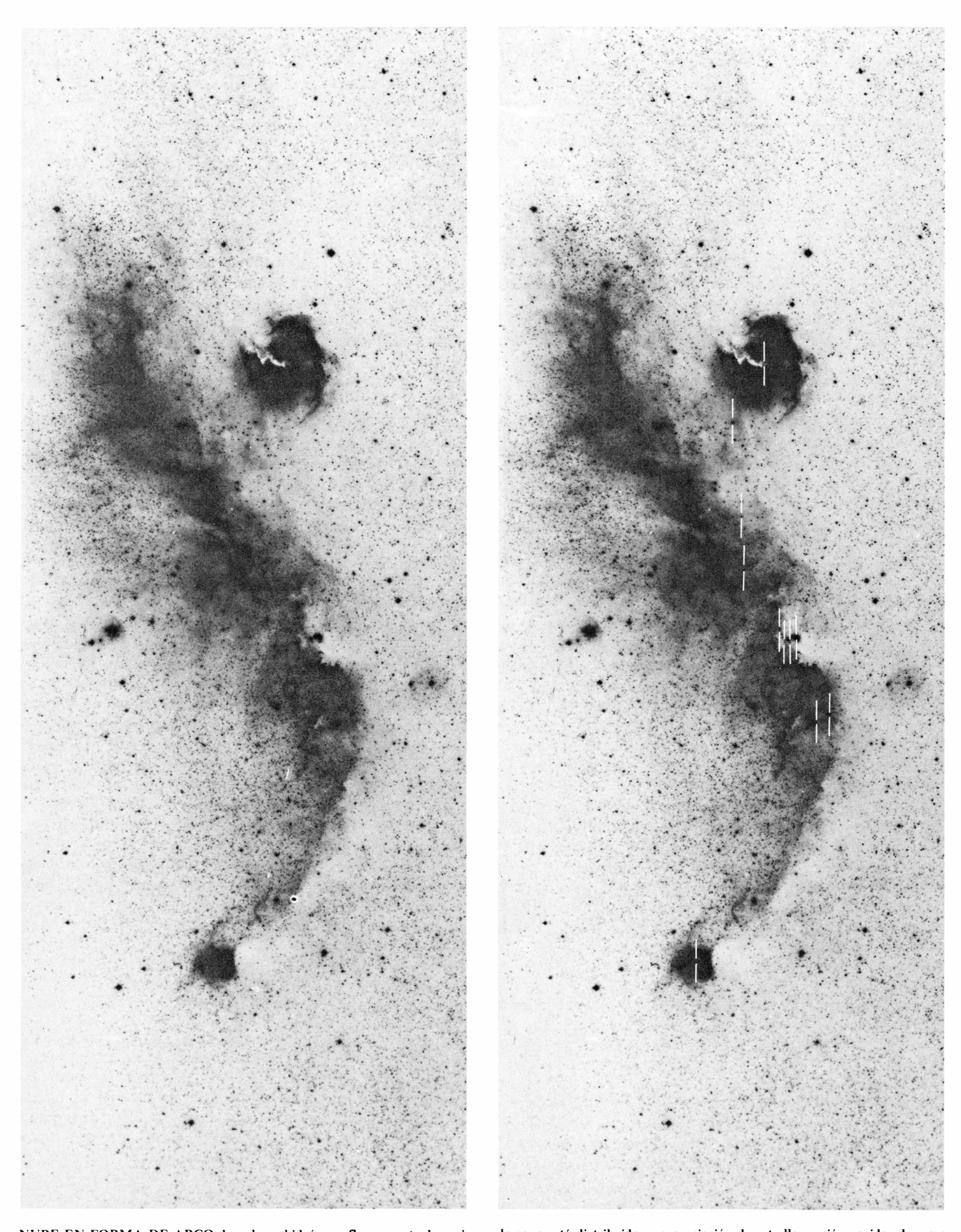
Hasta el pasado lustro faltaba confirmación observacional de la formación de estrellas inducida por las supernovas, por la sencilla razón de que los objetos que se habían identificado definitivamente como restos de supernovas eran demasiado jóvenes para asociarlos con el nacimiento de estrellas. Desde 1972, la situación ha cambiado en tres aspectos. En primer lugar, se han identificado capas de gas en expansión que rodean los restos de antiguas supernovas. Algunas de esas capas coinciden con enjambres de estrellas jóvenes. En segundo lugar, recientes avances en espectroscopía molecular y astronomía del infrarojo han permitido a los astrofísicos observar las fases primitivas de la formación de las estrellas: la compresión y calentamiento del medio interestelar. Tercero, el descubrimiento de concentraciones anómalas de ciertos isótopos en meteoritos indica que una supernova puede haber presidido el nacimiento del sistema solar.

El que una estrella se convierta o no en supernova depende de su masa. Cuando una nube ha alcanzado una densidad suficientemente alta, la mutua atracción gravitatoria entre sus partículas hace que se condense para formar una estrella. A medida que la materia se agrupa, las fuerzas gravitatorias se hacen más grandes, produciendo una mayor contracción de la estrella. La fuerza gravitatoria, en constante crecimiento, garantiza que la estrella continuará condensándose. A medida que la densidad de la estrella aumenta, también lo hace la temperatura de su interior. A temperaturas suficientemente altas, las reacciones termonucleares en el interior profundo de la estrella transmutan hidrógeno en helio. A partir de este punto, el curso del ciclo de vida de la estrella depende del equilibrio entre las fuerzas gravitatorias y la energía liberada por las reacciones de fusión, estando en realidad el colapso de la estrella impedido sólo temporalmente por las reacciones termonucleares.

El mecanismo exacto por el cual una estrella se convierte en supernova no se conoce todavía y puede incluso diferir de estrella a estrella. Uno de los mecanismos posibles es el siguiente. Después que la mayor parte del hidrógeno de su interior se ha convertido en helio, una estrella masiva se contrae hasta que su temperatura interior es suficiente para transmutar el helio en carbono. Cuando el helio se ha agotado, la estrella puede contraerse otra vez, y su temperatura aumenta constantemente hasta que el carbono empieza a su vez a transmutarse en elementos más pesados. En este punto puede surgir una supernova como resultado de la combustión explosiva del carbono o bien los elementos más pesados pueden fusionarse para formar otros todavía más pesados. Cada reacción termonuclear proporciona menos y menos energía, de modo que las fuerzas gravitatorias acaban ganando en la pugna para producir el colapso de la estrella.

¿Qué ocurre una vez que se ha agotado el combustible nuclear? Si la estrella es de menor masa que el Sol, se convierte en una enana blanca, estrella cuyo núcleo está formado de un gas altamente comprimido de núcleos atómicos (en su mayor parte núcleos de helio) y de los electrones arrancados de ellos. Si la estrella tiene una masa substancialmente más grande que la del Sol, las fuerzas gravitatorias autocomprimen la materia hasta un estado superdenso en el que los electrones y protones son empujados unos contra otros para formar neutrones. Los neutrones a su vez son forzados a soldarse de modo que formen un fluido nuclear. Si tal estrella de neutrones tiene una masa suficientemente grande, ni siquiera la extremadamente alta densidad de los neutrones estrechamente apretados puede detener el próximo colapso gravitatorio. Se cree que, al comenzar el colapso final, la estrella de neutrones se hace inestable y expulsa sus capas exteriores en una explosión de supernova. El núcleo de la estrella puede contraerse todavía más hasta que se convierte en un agujero negro, objeto cuyo campo gravitatorio es tan intenso que ni la materia ni la luz puede abandonarlo.

En nuestra galaxia se produce probablemente una supernova cada cincuenta años o así, aunque la mayoría de las supernovas se dan tan lejos que el polvo interestelar entre ellas y el sistema solar

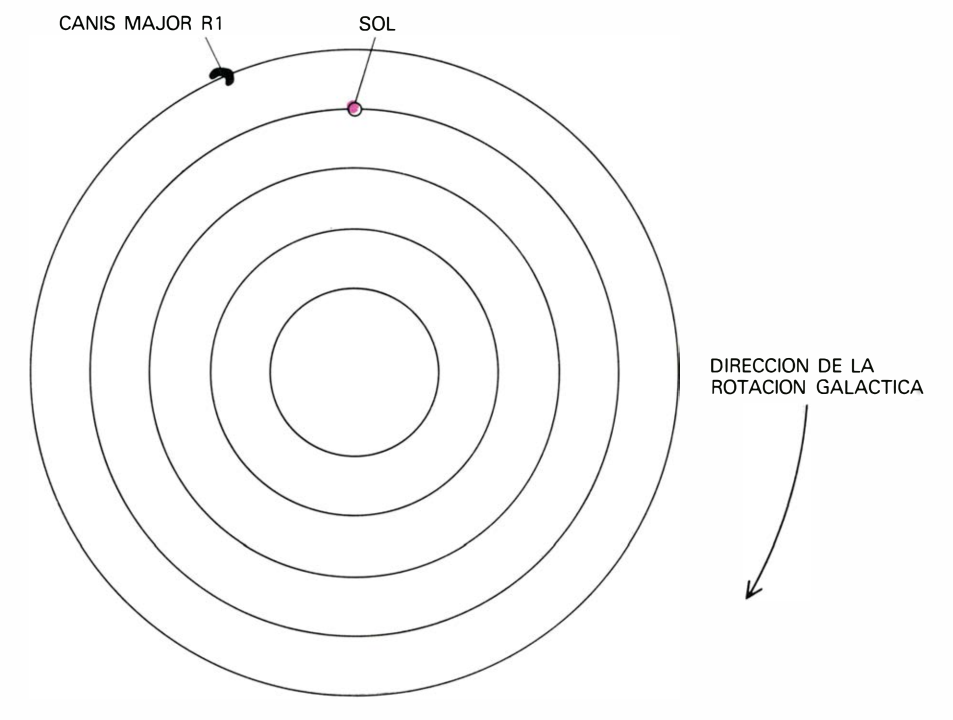


NUBE EN FORMA DE ARCO de polvo e hidrógeno fluorescente de casi 100 años-luz de arriba abajo, que aparece en estos dos negativos de una fotografía de una región en el Can Mayor hecha con el telescopio Schmidt de 122 centímetros del Monte Palomar. El arco es el borde de una capa esférica de gas en expansión. La energía de la expansión sugiere que la capa es lo que queda de una antigua supernova. A lo largo de la parte derecha de la capa está distribuida una asociación de estrellas recién nacidas de masa intermedia que se llama Canis Major R1. La capa puede haber desencadenado la formación de Canis Major R1 comprimiendo el gas y el polvo hasta una densidad suficientemente alta para que la gravedad reúna la materia formando estrellas. La luz reflejada de las estrellas jóvenes da lugar a nebulosas; la posición de algunas de éstas se marca a la derecha.

las hacen invisibles. Una supernova produce un efecto tal sobre el medio interestelar en su vecindad inmediata, que los restos de la explosión se pueden detectar durante al menos cien mil años. Las observaciones de restos de supernovas y las simulaciones de ellas realizadas con ordenador proporcionan un cuadro general de una explosión de supernova. La materia expulsada, que se mueve inicialmente con una velocidad de unos 10.000 kilómetros por segundo, choca con la materia enrarecida del medio interestelar y arrastra consigo parte de ella formando una capa en expansión, si bien decelerada. Al cabo de unos mil años, la materia ha viajado aproximadamente cuatro años-luz y se está dilatando a un ritmo de unos 1.000 kilómetros por segundo. Después de 50.000 años, los viejos restos de la supernova forman una capa gaseosa de un diámetro aproximado de 150 años-luz que se está expandiendo a un ritmo de unos 100 kilómetros por segundo. Los restos más antiguos que se pueden detectar comenzaron su viaje hacia el exterior hace unos 100.000 años; su diámetro es de unos 200 años-luz y su velocidad de expansión de unos 50 kilómetros por segundo.

Ciertamente, la evolución verdadera de un resto particular está regida por muchos factores, incluyendo la densidad y homogeneidad del medio ambiente interestelar y la cantidad precisa de energía liberada por la supernova. No obstante, el cuadro general que hemos descrito es una buena aproximación. Los residuos de antigüedad superior a 100.000 años se expanden con una velocidad de a lo más 50 kilómetros por segundo. Tales residuos pierden sus características distintivas de emisión óptica filiforme (radiación electromagnética emitida por filamentos de gas en los residuos) y radioemisión no térmica (con la intensidad de la radiación aumentando con la longitud de onda). Durante algún tiempo todavía pueden detectarse los restos en forma de capas de hidrógeno neutro (no ionizado) en expansión. Hacia el plano central de la galaxia, sin embargo, el hidrógeno es ya de por sí abundante, de modo que esas capas resultan difíciles de encontrar porque se funden con el medio interestelar. Por encima y por debajo del plano galáctico, el hidrógeno de fondo es más escaso. Carl E. Heiles, de la Universidad de California en Berkeley, ha demostrado que mucho del hidrógeno neutro que allí se encuentra está en forma de capas.

Se sabía desde hace tiempo que las supernovas hacían un papel de germen en el origen del sistema solar completamente aparte de la expansión de sus capas. Los elementos más pesados en las nubes de polvo y gas que se reunieron para formar el Sol y los planetas fueron expulsados por supernovas a lo largo de un período de varios miles de millones de años. (Los elementos más ligeros en las nubes, es decir, hidrógeno y helio, preexistían en su mayor parte.) El contenido de la materia expulsada difería probablemente de una supernova a otra, pero hacia el tiempo en que el sistema solar comenzó a formarse, los restos distintivos de miriadas de supernovas se habían entremezclado formando nubes de composición homogénea. En particular, la composición isotópica (es decir, la abundancia relativa de las especies nucleares de los elementos) de las nubes era homogénea. El reciente descubrimiento de meteoritos con una composición isotópica diferente de la media en el sistema solar significa que hubo al menos una supernova que explotó en época tan próxima a la de formación del sistema solar que su materia expulsada no se pudo mezclar completamente con la procedente de supernovas anteriores. Quizá la onda de choque de esta supernova desencadenó la formación del sistema solar [véase "¿Desencadenó una supernova la formación del sistema solar?" por David N. Schramm y Robert N. Clayton; INVESTIGACION Y CIENCIA, diciembre, 1978].



ROTACION DIFERENCIAL DE LA GALAXIA, la responsable de que el Sol se esté alejando de Canis Major R1. Las regiones interiores de la galaxia completan una vuelta alrededor del centro mucho más deprisa que las regiones exteriores; de aquí que el Sol tenga un período orbital de 250 millones de años, mientras que Canis Major R1 tiene un período de 270 millones de años. El abundante hidrógeno existente entre Canis Major R1 y el Sol, y por tanto entre Canis Major R1 y el sistema solar, parece estar alejándose a causa de la velocidad diferencial de rotación. La frecuencia a la cual la radiación emitida por el hidrógeno (1.420.4057 megahertz) se recibe en la Tierra ha sufrido un corrimiento Doppler consecuencia del movimiento del hidrógeno respecto al receptor. Como el hidrógeno se está alejando, el corrimiento se produce hacia frecuencias más bajas. Cerca del foco del arco de Canis Major R1, sin embargo, se está emitiendo radiación que ha sufrido corrimiento Doppler hacia frecuencias más altas. Esto indica que el hidrógeno de allí, que constituye la capa en expansión que queda de la supernova, está acercándose hacia la Tierra a velocidades de a lo más 30 kilómetros por segundo.

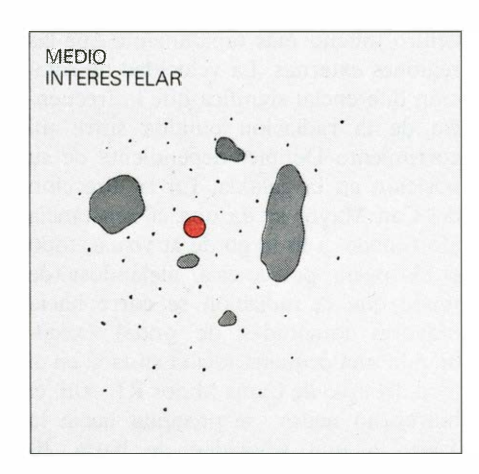
El sistema solar tiene aproximadamente sólo un tercio de la edad de la galaxia, de modo que es razonable admitir que las condiciones que conducían a la formación de estrellas en el tiempo en que se formó el Sol son las mismas que actúan hoy. No hay duda de que se han formado recientemente estrellas en la vecindad del Sol y continúan formándose. Por ejemplo, la Gran Nebulosa de Orión es un agregado de estrellas luminosas de gran masa a una distancia de unos 1500 años-luz. Estas estrellas de gran masa son blanco-azuladas de los tipos espectrales O y B. Tienen temperaturas superficiales entre los 16.000 y los 45.000 grados Kelvin, comparados con los 6000 del Sol, y son entre 800 y 500.000 veces más luminosas que éste. Tales agregados, llamados asociaciones OB, constan de entre 5 y 50 estrellas

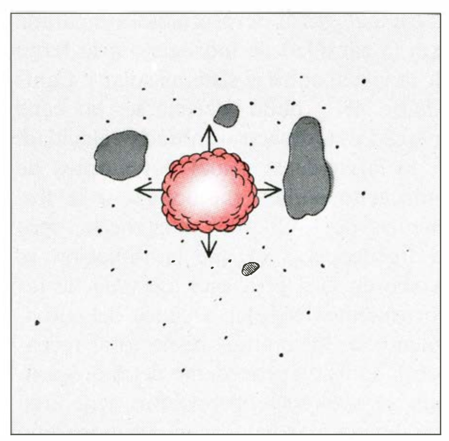
cada uno, con una masa de aproximadamente entre 10 y 30 veces la masa del Sol. Las estrellas del tipo O y tipo B emiten energía a un ritmo tan prodigioso que no pueden sobrevivir por más de 10 millones de años (probablemente una milésima parte de la edad de la galaxia), de modo que se tienen que haber formado en época relativamente reciente.

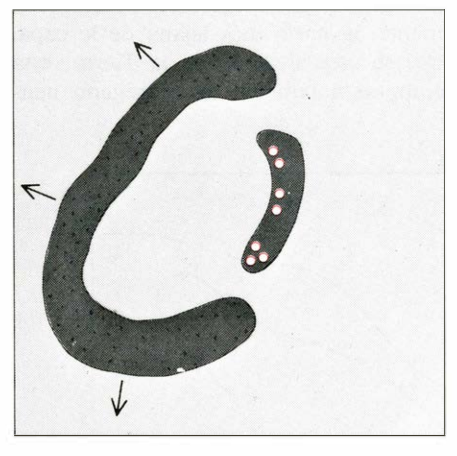
Dispersas entre las estrellas de la asociación OB de la nebulosa de Orión hay estrellas menos luminosas que tienen entre 100.000 y 10 millones de años de edad. Tales estrellas, con masas inferiores a tres veces la del Sol, son más corrientes que los tipos espectrales O y B. Las estrellas con esas masas relativamente bajas tardan más en formarse (ya que las fuerzas gravitatorias que tienden a reunirlas son más débiles) y no han evolucionado todavía hasta la fase estable de combustión de hidrógeno en la que comienza la síntesis de los elementos más pesados. Se llaman pre-fase principal. Las que muestran líneas de emisión en sus espectros se llaman estrellas T Tauri, por estar su prototipo en la constelación de Tauro. Muchas asociaciones T (agregados de estrellas T Tauri) están situadas dentro de los límites de las asociaciones OB, aunque algunas se encuentran en regiones en donde no existen estrellas masivas.

En 1966, Sydney van den Bergh, de la Universidad de Toronto, descubrió una nueva clase de agregado estelar: las asociaciones R, muchas de las cuales sirven de lugar de nacimiento de estrellas de masa intermedia (entre tres y 10 masas solares). Encontró las asociaciones R examinando fotografías de la Vía Láctea tomadas como parte del programa de exploración del cielo del Observatorio Palomar. Buscaba estrellas que fuesen tan próximas a las nubes de polvo interestelar que la luz reflejada por las nubes se viese en forma de nebulosas. Probablemente, en las asociaciones R están también formándose estrellas cuya masa es inferior a tres masas solares, pero son demasiado débiles para crear nebulosas de reflexión y por tanto deben identificarse mediante otros procedimientos. Existen muchas asociaciones R dentro de los límites de asociaciones OB.

Una asociación R de considerable interés es la que está situada en la constelación del Can Mayor, sólo a unos pocos grados de Sirio. La asociación, Canis Major R1, consta de unas 30 estrellas en nebulosas de reflexión que están dispuestas principalmente en una nube de polvo en forma de arco con una extensión de unos 100 años-luz. También hay aquí una asociación OB, Canis Major OB1, cuyas estrellas se extienden más allá de la nube de polvo. La distancia entre el sistema solar y Canis Major R1 es de unos 3700 años-luz.









FORMACION DE ESTRELLAS inducida por una supernova. Comienza con una estrella de gran masa (*en color*) inmersa en el medio interestelar cargado de nubes (*parte superior izquierda*). Al contraerse la estrella bajo la atracción gravitatoria de las partículas que la constituyen, se hace inestable y expulsa sus capas más externas en una explosión de supernova. Después de unos 30.000 años, los restos tienen 100 años-luz de anchura y han barrido una capa de materia interestelar que se está expansionando a una velocidad de al menos 100 kilómetros por segundo (*parte superior derecha*). Después de 100.000 años, los restos tienen un diámetro de 200 años-luz y su velocidad ha descendido hasta a lo más 50 kilómetros por segundo (*parte inferior izquierda*). La onda de choque de la capa puede haber comprimido las nubes de gas y polvo hasta una densidad suficientemente alta para que la gravedad comience a reunir la materia para formar estrellas (*puntos coloreados*). Después de unos tres millones de años, la asociación está bien desarrollada (*parte inferior derecha*). Por esta época se han formado muchas estrellas que se arraciman en un grupo, rodeado por un gran lazo de hidrógeno en lenta expansión. El proceso completo puede repetirse si una de las estrellas recién nacidas llega a desarrollarse hasta la fase de supernova (catastrófica explosión de una estrella al final de su vida.)

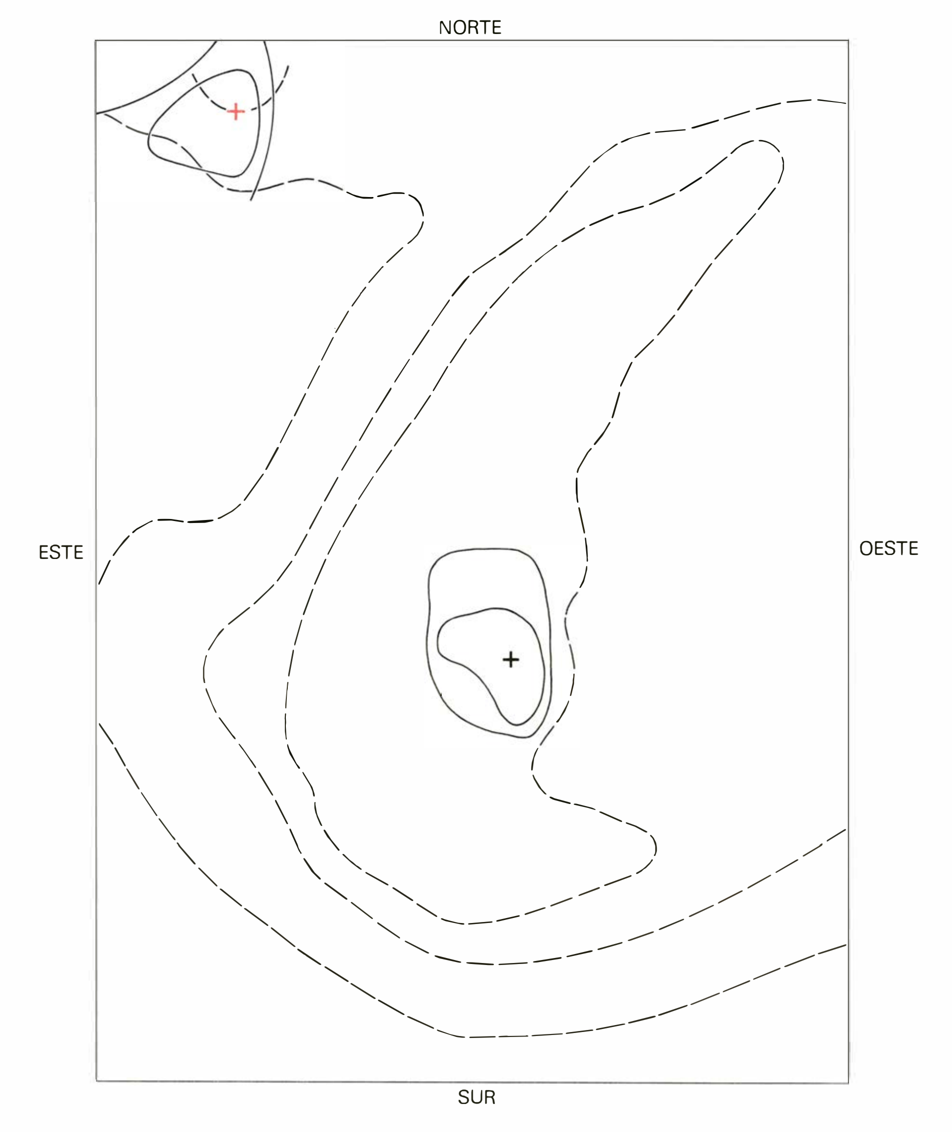
La asociación R nos intriga por dos razones. En primer lugar, incluye estrellas de masa intermedia que todavía parecen estar en la fase pre-principal de su ciclo de vida. De acuerdo con un estudio realizado por uno de nosotros (Herbst), René Racine, de la Universidad de Montreal y John W. Warner, de la Universidad de Minnesota, tales estrellas tienen probablemente unos 300.000 años de edad, lo que las sitúa entre los objetos estelares más jóvenes que se conocen. En segundo lugar, la nube de polvo en forma de arco en la que las estrellas se están formando está situada a lo largo de uno de los bordes de una capa gigante de gas en expansión.

Esta capa fue identificada por primera vez a partir de las radioondas de 21 centímetros emitidas por el hidrógeno neutro en la capa. Un átomo de hidrógeno está constituido por un protón y un electrón, que pueden girar bien en el mismo sentido o en sentidos opuestos. Si los giros tienen lugar en el mismo sentido, el electrón puede voltear de tal manera que su giro sea opuesto al del protón. Cuando el electrón voltea de esta manera, el átomo de hidrógeno emite radiación a una frecuencia calculada de 1.420.4057 megahertz, lo que corresponde a una longitud de onda de unos 21 centímetros, en la banda de las radioondas.

Aunque la cantidad de hidrógeno en

la capa de gas es despreciable comparada con la cantidad de hidrógeno a lo largo de la visual entre el sistema solar y Canis Major R1, pudo detectarse la capa merced a su desacostumbrada velocidad. A lo largo de la visual, los átomos de hidrógeno emiten radioondas a la frecuencia de 1.420.4057 megahertz, pero la frecuencia a la que la radiación se recibe en la Tierra está afectada de un corrimiento Doppler a causa del movimiento de los átomos respecto al receptor: la emisión procedente del hidrógeno que se aleja del observador tiene una frecuencia más baja, y, la del hidrógeno que se acerca, una frecuencia más alta.

La galaxia es un cuerpo dotado de rotación diferencial: sus regiones interiores completan un circuito alrededor del centro mucho más rápidamente que las regiones externas. La velocidad de rotación diferencial significa que la frecuencia de la radiación emitida sufre un corrimiento Doppler dependiente de su posición en la galaxia. En la dirección del Can Mayor se da una circunstancia afortunada: a lo largo de la visual, todo el hidrógeno parece estar alejándose (de modo que la radiación se corre hacia mayores longitudes de onda) exceptuando una pequeña región situada en el foco del arco de Canis Major R1. Allí, el hidrógeno neutro se precipita hacia la Tierra a una velocidad de hasta 30 kilómetros por segundo. Presumiblemente, la parte más lejana de la capa, que se está alejando de la Tierra, está compuesta también de hidrógeno neutro. Es imposible detectar la parte más alejada porque, a lo largo de la visual, todo el hidrógeno de fondo se mueve alejándose asimismo.



DOS FUENTES DE RADIACION INFRARROJA (*signos más*) inmersas en nubes moleculares (*líneas continuas*) asociadas con los antiguos restos de una supernova W44 (*líneas de trazos*). Las ondas infrarrojas son emitidas por envolturas de polvo que rodean las estrellas en su fase inicial de evolución. El polvo absorbe radiación de las estrellas, se calienta y reemite energía a longitudes de onda en el infrarrojo. Una de las dos fuentes infrarrojas (*signo más en color*) está decididamente asociada con W44, mientras que es posible que la otra fuente (*signo más en negro*) sea sólo un objeto de fondo. Ambas fuentes infrarrojas fueron descubiertas por H. Alwyn Wooten, de la Universidad de Texas. El mapa cubre alrededor de 30 minutos de arco en el cielo, que es equivalente al diámetro de la luna llena.

Ronald J. Reynolds y Peter M. Ogden, de la Universidad de Wisconsin, estudiaron intensivamente la cinemática del gas en la vecindad de Canis Major R1. Las fotografías tomadas como parte del proyecto de exploración del cielo del Observatorio Palomar habían revelado una distribución de gas ionizado en forma de anillo adyacente a la asociación R. Con un espectrómetro Fabry-Pérot, Reynolds y Ogden observaron las líneas de emisión de longitudes de onda en el visible en una región ionizada del gas. Las líneas de emisión resultan de la radiación de energía por los átomos cuando pasan de órbitas de mayor energía a otras de energía más baja. En el laboratorio terrestre, las transiciones entre ciertas órbitas en ciertos átomos están "prohibidas"; resultan extremadamente improbables porque es mucho más verosímil que los átomos pierdan su exceso de energía en choques con átomos vecinos. La densidad de la materia en el espacio interestelar es tan baja, sin embargo, que los choques entre los átomos son escasos y muy espaciados. De aquí que los átomos algunas veces radien en líneas prohibidas.

Reynolds y Ogden detectaron una línea prohibida del nitrógeno hacia el centro del anillo de gas ionizado. La línea era doble, es decir, constaba de dos frecuencias porque había corrimiento Doppler en dos sentidos. Esto indicaba que el gas del centro del anillo estaba dividido en dos regiones cuyas velocidades diferían en 26 kilómetros por segundo. Cerca del borde del anillo, la línea no estaba duplicada. Estas observaciones confirman la presencia de una capa de gas en expansión adyacente a la asociación R; las regiones de velocidades opuestas constituyen la parte anterior y posterior de la capa.

¿Cuál es el origen de la capa de gas en expansión? En nuestra opinión, hacer esta pregunta equivale a preguntar por qué razón se forman estrellas en Canis Major R1, dado que es extremadamente improbable que las estrellas se hayan formado por casualidad a lo largo de un arco de 100 años-luz dentro de una capa en expansión de 200 años-luz de diámetro. Como hemos visto, una explosión de supernova genera la energía suficiente para producir tal capa.

Basándonos en la hipótesis de que la capa de Canis Major R1 es el resto de

una supernova, calculamos su edad a partir del diámetro y velocidad de expansión. Encontramos que la capa tenía una edad de 800.000 años. Como las estrellas en la asociación R tienen una edad de unos 300.000 años, la edad de las estrellas concuerda con nuestra hipótesis de que una supernova desencadenó su nacimiento. La densidad de materia en la capa en expansión se puede estimar a partir de la cantidad de hidrógeno ionizado que rodea a tres estrellas dentro de la capa. A partir de la densidad, el diámetro y la velocidad de expansión, pudimos calcular la cantidad de energía que se había necesitado para crear la capa: aproximadamente 10^{44} joule, comparable a la energía liberada en una explosión de supernova.

¿Puede algún otro fenómeno liberar una tal cantidad de energía? La respuesta es afirmativa. Una estrella de gran masa podría inyectar 10^{44} joule en una pequeña región del espacio en el curso de su ciclo de vida de varios millones de años. No se sabe, sin embargo, si los productos expulsados por una estrella masiva podrían comprimir en forma efectiva la materia del medio interestelar. Los modelos teóricos sugieren que en un medio homogéneo, las estrellas de tipo O podrían dar lugar a capas en expansión capaces de desencadenar la formación de estrellas. En realidad, el medio interestelar dista mucho de ser homogéneo, de modo que los modelos no pueden aplicarse. Además, no hay ninguna estrella luminosa de gran masa cerca del centro de la capa de Canis Major R1 que hubiera probablemente podido servir como fuente de energía. Las dos estrellas de mayor masa en la vecindad del anillo están situadas a lo largo de su borde, no cerca de su centro. En las mejores fotografías de Canis Major R1 se observan estructuras débiles y parcialmente en forma de anillo que rodean las dos estrellas, pero las estructuras son insignificantes comparadas con el anillo de Canis Major.

Una de esas dos estrellas de gran masa es particularmente interesante porque es una estrella errante: una estrella cuya velocidad, de entre 30 y 200 kilómetros por segundo, es muy grande comparada con las velocidades de las estrellas vecinas. Adriaan Blaauw, de la Universidad de Leiden, ha desarrollado la explicación más ampliamente aceptada acerca de las estrellas errantes. La mayoría de las estrellas nacen como miembros de un sistema binario: dos estrellas que giran alrededor de su centro de masa común como consecuencia de su atracción gravitatoria mutua. La de mayor masa de un sistema binario evoluciona más deprisa y puede convertirse con el tiempo en una supernova. En ese momento, el sistema binario expulsa instantáneamente una gran cantidad de masa y la estrella menos evolucionada adquiere una mayor velocidad con respecto a las estrellas vecinas.

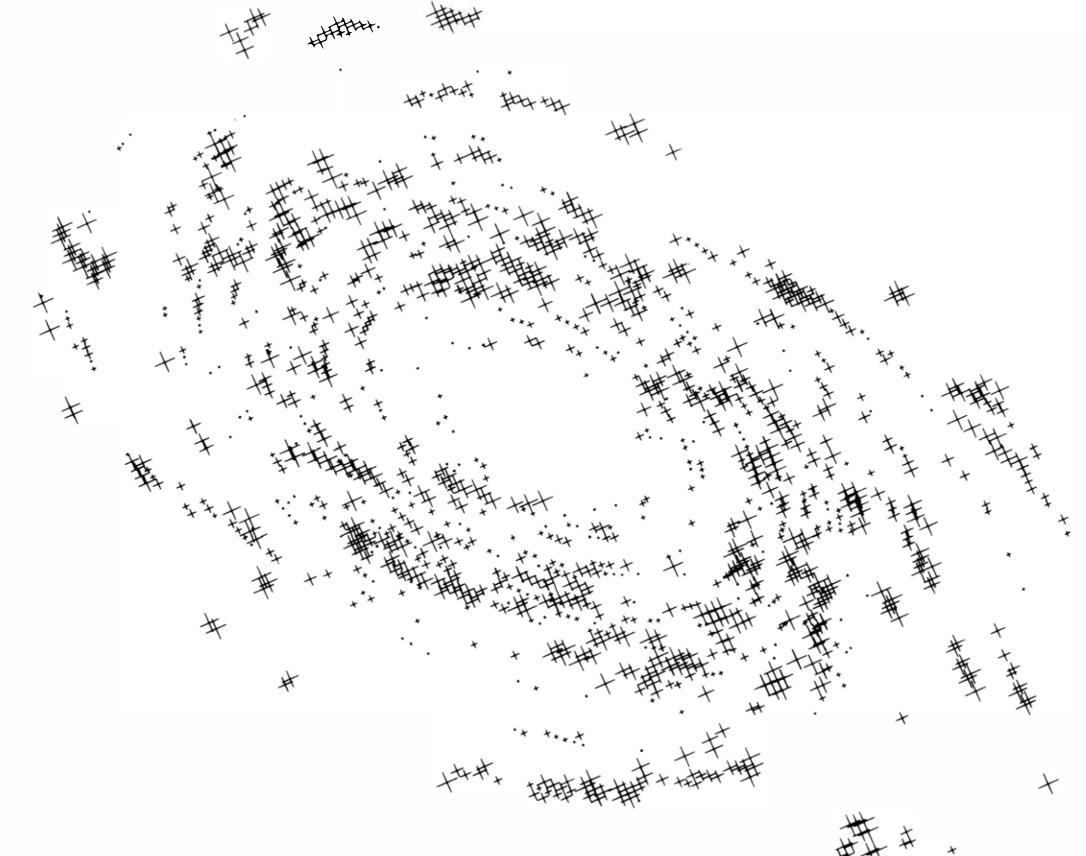
La errante situada cerca de Canis Major R1 se está alejando del Sol con una velocidad 30 kilómetros por segundo mayor que las velocidades de las estrellas en su vecindad. Como la mayor parte de las errantes, ésta es caliente y luminosa y viaja sola. Hasta ahora sólo se ha determinado la componente de la velocidad según la visual, de modo que todavía no es posible decir si la errante estuvo tiempo atrás en el centro de la capa en expansión. Sin embargo, si la velocidad tangencial (o perpendicular a la visual) es comparable a la velocidad según ésta, la estrella puede muy bien haber estado en el centro de la capa hace unos 800.000 años, cuando la capa se formó.

Otro hecho adicional relaciona la errante con la formación de la capa. En su estudio espectroscópico, Reynolds y Ogden encontraron que una línea de emisión del oxígeno se refuerza a velocidades que son características de la parte exterior de la capa. Este descubrimiento se hace comprensible si la errante está en la actualidad dentro o cerca del borde exterior de la capa, y ahí es exactamente donde debería estar si fuera la compañera original de la supernova que produjo la capa y si adquirió su velocidad de alejamiento de 30 kilómetros por segundo en el momento de la explosión.

El Canis Major R1 es un ejemplo claro de que las estrellas se forman en el borde de una capa en expansión, pero no es en manera alguna el único. Monoceros R1, una asociación R a unos 2500 años-luz del Sol, ha sido examinada recientemente por Marc L. Kutner, Robert L. Dickman y sus colaboradores en el Instituto Politécnico Rensselaer. Las estrellas en Monoceros R1 se parecen a las de Canis Major R1 y tienen probablemente casi la misma edad. Como Canis Major R1, Monoceros R1 consta de estrellas recién nacidas situadas a lo largo de una distribución en forma de anillo de nubes densas de gas y polvo interestelar. El diámetro del anillo mide unos 25 años-luz, 10 veces menos que el de Canis Major R1. La diferencia sugiere que la densidad inicial de la materia que formó el anillo era mayor en Monoceros R1. Hay pruebas de la existencia de hidrógeno neutro en expansión cerca del centro de este anillo, también, y el ritmo de energía de expansión es comparable al de Canis Major R1. En Monoceros R1 no hay estrellas de gran masa, de modo que no hay anillos parciales de gas luminoso ionizado como los de Canis Major. Todas las indicaciones apuntan a la conclusión de que en Monoceros R1 explotó una supernova. La única posible fuente de energía distinta, una estrella de masa extremadamente grande, está ausente.

En dos nubes en el borde de la Nebulosa Gum, en el cielo del Hemisferio Sur, Richard D. Schwartz, de la Universidad de Missouri, descubrió objetos Herbig-Haro, nebulosas semiestelares asociadas de una manera poco conocida con las fases primitivas de la formación de estrellas. Argüía que la expansión de la Nebulosa Gum, que pudo deber su origen a una supernova al menos, ha desencadenado la formación de estrellas en estas nubes. Hay indicaciones circunstanciales en apoyo de esta hipótesis. Cerca de la Nebulosa Gum, T. G. Hawarden y P. W. J. L. Brand hallaron varios "glóbulos cometarios": nubes de polvo con colas alargadas que son compactas, bastante opacas, aisladas y de forma regular. Sin excepción, las colas alargadas apuntan hacia fuera del centro de la nebulosa Gum. La nebulosa NGC 5367 está en un glóbulo cometario en el Centauro que coincide con un cúmulo de jóvenes estrellas de pequeña masa, según Howard Van Till y sus colaboradores en la Universidad de Texas y P. Williams y sus colaboradores de la Universidad de Edimburgo. Quizá los glóbulos cometarios de la Nebulosa Gum son sedes de formación de estrellas o podrían llegar a serlo en el futuro.

Hasta ahora hemos limitado la explicación a las estrellas jóvenes, aunque visibles, y a los restos de supernovas tan antiguos que sólo hay indicaciones indirectas de que sean realmente restos de supernovas y no otra cosa. Algunas indicaciones de la formación de estrellas inducida por las supernovas provienen también del descubrimiento de nubes moleculares que están evolucionando hacia estrellas en la proximidad de restos conocidos de supernovas. Por ejemplo, H. Alwyn Wooten, de la Universidad de Texas, ha cartografiado una nube comprimida situada en el borde de un antiguo resto de supernova conocido como Lazo del Monoceros y ha encontrado condiciones susceptibles de conducir a la formación de estrellas. En nubes asociadas con otro resto, el W 44, descubrió al menos una fuente de radiación



GALAXIA ESPIRAL M81 (*arriba*). Quizá debe su aparición a una supernova. Una teoría de la estructura de las galaxias espirales propuesta por Huberto Gerola y Philip E. Seiden, del Thomas J. Watson de IBM, sostiene que si una supernova desencadenara la formación de estrellas capaces de originar otra supernova, se podría entonces producir una cadena de regiones de formación de estrellas. Si se creasen muchas de tales cadenas en una galaxia en rotación diferencial, la distribución de estrellas formaría una estructura espiral. Simulaciones de esta teoría realizadas con ordenador para la espiral M81 (*abajo*) dan origen a una distribución de brillantes estrellas recién nacidas (*cruces*) que se parece a la distribución real (*arriba*).

infrarroja y posiblemente dos. Se cree que tales fuentes infrarrojas indican la presencia de estrellas jóvenes. La radiación es emitida por las envolturas de polvo que desprenden las estrellas en la fase inicial de su evolución. El polvo absorbe la radiación de las estrellas, se calienta y a su vez reemite en longitudes de onda del infrarrojo.

Es el momento de considerar hasta qué punto está extendida la formación de estrellas inducida por supernovas. La mayoría de las estrellas parecen haberse formado en asociaciones, que a veces están compuestas de subgrupos separados de estrellas, las más jóvenes de las cuales se hallan envueltas en gas y polvo. Blaauw ha sugerido que Orión OB1 consta de cuatro subgrupos alineados en orden de edades (del subgrupo más antiguo en un extremo de la asociación al más moderno en el otro extremo), de modo que la asociación se puede haber creado en brotes sucesivos de formación de estrellas.

Hay un modelo de la naturaleza explosiva de la formación de estrellas en una asociación OB que no hace uso de las supernovas. Este modelo detallado, desarrollado por Bruce G. Elmegreen y Charles J. Lada, del Centro de Astrofísica del Observatorio del Harvard College y el Observatorio Astrofísico Smithsoniano, se centra sobre la copiosa cantidad de radiación ultravioleta emitida por las estrellas recién nacidas, que ioniza el hidrógeno en el medio ambiente interestelar. La región de hidrógeno ionizado comprime una nube molecular próxima, creando una onda de choque. Cuando la materia en el borde de la nube se hace suficientemente densa, se rompe en fragmentos que se reúnen por efecto gravitatorio formando las estrellas de un subgrupo. Con el tiempo, estas estrellas crean una nueva región de hidrógeno ionizado, que genera una onda de choque que desencadena el nacimiento de un nuevo subgrupo. El proceso se repite de nuevo hasta que se ha formado la asociación completa. Por supuesto, este modelo no explica cómo se formó el subgrupo inicial de estrellas recién nacidas; la presión necesaria para ello podría haber sido suministrada por una onda de choque de otra nube molecular o con una explosión de supernova.

Aunque las supernovas no se han incorporado todavía en un modelo detallado de la naturaleza explosiva de una asociación OB, pueden desempeñar un papel determinante en el proceso. Si una supernova desencadenase un brote de formación de estrellas en una densa nube primitiva, y si la estrella de mayor masa que se formase de la nube pudiera evolucionar hasta la fase de supernova y desencadenase una nueva serie de formación de estrellas que a su vez produjese otra supernova, y así sucesivamente, se explicaría entonces la distribución de estrellas observada en Orión OB1. Además, el tiempo transcurrido entre brotes sucesivos sería de varios millones de años, idéntico a la diferencia de edad observada entre los distintos subgrupos de estrellas.

Independientemente de que las supernovas sean o no responsables de la estructura en subgrupos de Orión OB1, no hay duda de que con frecuencia se encuentran dentro de las asociaciones. Blaauw fue el primero en darse cuenta de este hecho cuando se propuso averiguar la historia de tres estrellas errantes hasta su origen en Orión OB1. Observó también que en los subgrupos más antiguos en una asociación hay menos estrellas extremadamente masivas que en los subgrupos más jóvenes. Contando el número de esas "estrellas que faltan", se encuentra que en los pasados 10 millones de años explotaron en Orión OB1 alrededor de 20 supernovas. La enorme capa que esas supernovas han generado colectivamente fue estudiada por Reynolds y Ogden, quienes estiman que su contenido de energía es de unos 10^{45} joule. La última supernova en Orión probablemente data de hace 500.000 años, y puede ser el origen de un frente de choque observado recientemente por Lennox L. Cowie y sus colaboradores de la Universidad de Princeton.

Si una supernova desencadenase la formación de estrellas que diera lugar a otra supernova, resultaría entonces una cadena de regiones sede de formación de estrellas. Si se creasen muchas de tales cadenas en una galaxia en rotación diferencial, la distribución de estrellas se parecería a la observada en una galaxia espiral. Esa posibilidad es la base de un nuevo modelo de galaxia espiral propuesto por Huberto Gerola y Philip E. Seiden, del Centro de Investigación Thomas J. Watson de la International Business Machines Corporation. La simulación realizada con ordenador de la formación de estrellas que se autopropaga en discos giratorios en rotación diferencial pudo reproducir la apariencia de muchas galaxias espirales sin requerir la presencia de una onda de densidad concomitante. Tal onda de densidad es el rasgo más saliente de la teoría de mayor aceptación sobre la estructura en gran escala de las galaxias, desarrollada por C. C. Lin, del Instituto de Tecnología de Massachusetts, y Frank H. Shu de la Universidad de California en Berkeley. La teoría mantiene que una onda de densidad de forma espiral barre el plano central de la galaxia, comprimiendo nubes de gas y polvo hasta cinco o 10 veces su densidad original. Las nubes se contraen hasta formar estrellas dispuestas en una configuración espiral. Se está ahora trabajando para definir con precisión las diferencias entre las estructuras galácticas predichas por la teoría de la onda de densidad y las predichas por el modelo de supernovas. Las estructuras predichas se compararán entonces con las observadas.

Es posible que el universo sea suficientemente diverso para que ambas teorías estén en lo cierto. Algunas galaxias tienen dos bellos brazos que, partiendo del núcleo galáctico, se dirigen en espiral hacia fuera simétricamente. Esta clase de estructura no puede explicarse fácilmente con el modelo de supernovas. Otras galaxias parecen no tener una estructura espiral coherente en gran escala; están formadas enteramente de trozos de brazos espirales distribuidos en una forma al parecer caótica. Tales galaxias no pueden explicarse fácilmente mediante el modelo de onda de densidad. Quizá cada uno de esos mecanismos contribuye a la estructura espiral de una galaxia típica.

Nuestra propia galaxia puede ser un buen ejemplo. Desde hace 25 años se sabe que el Sol está situado dentro de una larga y estrecha concentración de brillantes estrellas que se parece a uno de los brazos espirales de otras galaxias. La concentración se conoce por Brazo Local. Se han detectado secciones de otros dos brazos espirales, pero no se puede cartografiar la estructura completa de la galaxia a longitudes de onda dentro del visible porque el polvo en el plano central de la galaxia entorpece las observaciones. Hay indicaciones de que los dos brazos parcialmente identificados se deben a ondas de densidad porque parecen ejercer una atracción gravitatoria sobre las estrellas y gas en la vecindad. No hay indicios de que el Brazo Local ejerza tal fuerza, según Lin y sus colaboradores. La teoría de la onda de densidad predice que el Sol debería estar casi exactamente entre brazos espirales, mientras que en realidad está situado en una región de activa formación de estrellas. Esto podría explicarse si las supernovas fuesen las principales responsables de la formación de estrellas en el Brazo Local.

Problemas físicos con muchas escalas de longitud

Sistemas físicos tan distintos como imanes y fluidos se parecen en que tienen fluctuaciones de estructura a lo largo de un amplio rango de tamaños. Se ha inventado un nuevo método para explicarlos llamado grupo de renormalización

Kenneth G. Wilson

De las propiedades de la naturaleza, una de las más prominentes es la gran diversidad de tamaños o escalas de longitud en la estructura del mundo. Un océano, por ejemplo, tiene corrientes que se mantienen durante miles de kilómetros y mareas que implican a todo el globo; el tamaño de sus olas puede oscilar desde menos de un centímetro a varios metros. Para una resolución mas fina, debe considerarse el agua de mar como un agregado de moléculas cuya escala de longitud característica es aproximadamente 10^{-8} centímetros. De la estructura menor a la mayor hay un abanico de unos diecisiete órdenes de magnitud.

En general, los sucesos que se diferencian por una gran disparidad de tamaño tienen poca influencia mutua; no hay nada en común entre ellos, y, por tanto, los fenómenos asociados con cada escala pueden tratarse independientemente. La interacción entre dos moléculas de agua vecinas es la misma si las moléculas están en el Océano Pacífico que si están en una taza de té. E igualmente importante es que una ola del océano pueda describirse con bastante precisión como una perturbación de un fluido continuo, sin atender en absoluto a la estructura molecular del líquido. El éxito de casi todas las teorías usuales en física se basa en aislar un rango limitado de escalas de longitud. Si, en las ecuaciones de la hidrodinámica, hubiera que especificar el movimiento de cada molécula de agua, no se alcanzaría, con los medios de que dispone la ciencia hoy, ninguna teoría sobre la dinámica de las olas del océano.

Pero existe una clase de fenómenos, en cuya explicación concurren, con igual importancia, sucesos de muchas escalas de longitud. Tenemos un ejemplo en el comportamiento del agua cuando se calienta hasta su ebullición a una presión de 217 atmósferas. A esa presión, el agua sólo empieza a hervir cuando la temperatura alcanza los 647 grados Kelvin. Esta combinación de presión y temperatura define el punto crítico del agua, donde desaparece la distinción entre líquido y gas; a presiones mayores hay una única fase fluida indiferenciada, y el agua no puede hacerse hervir por mucho que se eleve la temperatura. Cerca del punto crítico el agua genera fluctuaciones de su densidad en todas las escalas posibles. Las fluctuaciones toman la forma de gotas de líquido completamente mezcladas con burbujas de gas, y hay tanto gotas como burbujas de todos los tamaños, desde simples moléculas hasta el volumen de la muestra. En el punto crítico, precisamente, la escala de las mayores fluctuaciones se hace infinita, sin que disminuyan las fluctuaciones menores. La teoría que pretenda describir el agua en la proximidad de su punto crítico debe considerar todo el espectro de escalas de longitud.

La multiplicidad de escalas de longitud complica muchos de los problemas no resueltos de la física teórica y de algunos otros campos de estudio. Se han encontrado soluciones exactas para tan sólo unos cuantos problemas; para otros, hasta las mejores aproximaciones resultan insatisfactorias. En la pasada década se introdujo un nuevo método llamado grupo de renormalización para abordar problemas en los que concurren muchas escalas de longitud. No es que con él los problemas pierdan su dificultad, pero algunos que han resistido el tratamiento con otros enfoques pueden ceder ante este método.

El grupo de renormalización no es una teoría descriptiva de la naturaleza, sino un método general de construir teorías. Puede aplicarse no sólo a un fluido en el punto crítico sino también a un material ferromagnético a la temperatura en que se establece la magnetización o imanación espontánea, o a una mezcla de líquidos a la temperatura en que se hacen totalmente miscibles, o a una aleación a la temperatura a la que dos clases de átomos metálicos se distribuyen ordenadamente. Otros problemas apropiados incluyen el flujo turbulento, el establecimiento de la superconductividad y de la superfluidez, la formación de polímeros y el enlace de las partículas elementales llamadas quarks. Una hipótesis importante que parece confirmarse con los trabajos basados en el grupo de renormalización es que algunos de estos fenómenos, que considerados en su superficie se dirían muy distintos, son idénticos en un nivel más profundo. Así, el comportamiento crítico de los fluidos, los ferromagnetos o ferroimanes, las mezclas líquidas y las aleaciones pueden describirse mediante una teoría única.

El marco más conveniente para discutir cómo funciona el grupo de renormali-

MULTIPLES ESCALAS DE LONGITUD caracterizan las configuraciones que aparecen cuando un sólido ferromagnético se enfría hasta la temperatura en que se magnetiza espontáneamente. Cada cuadro representa el momento magnético asociado con un solo átomo del sólido; se supone que cada momento tiene sólo dos orientaciones posibles, denominadas "arriba" (*cuadrados negros*) y "abajo" (*cuadrados blancos*). A alta temperatura (*parte superior*), la orientación de los momentos magnéticos es esencialmente al azar, de manera que sólo hay una ordenación de corto alcance en la configuración. Cuando la temperatura se reduce (*medio*) empiezan a desarrollarse zonas algo mayores en las que la mayoría de momentos magnéticos están alineados en la misma dirección. Cuando la temperatura alcanza un valor crítico, llamado la temperatura de Curie, o T_c (*parte inferior*), estas zonas crecen hasta un tamaño infinito; sin embargo, significativamente, las fluctuaciones de menor escala persisten. Como resultado deben incluirse todas las escalas de longitud en una descripción teórica del ferroimán. Esta simulación de un ferroimán fue realizada, con un ordenador, por Stephen Shenker y Jan Tobochnik.

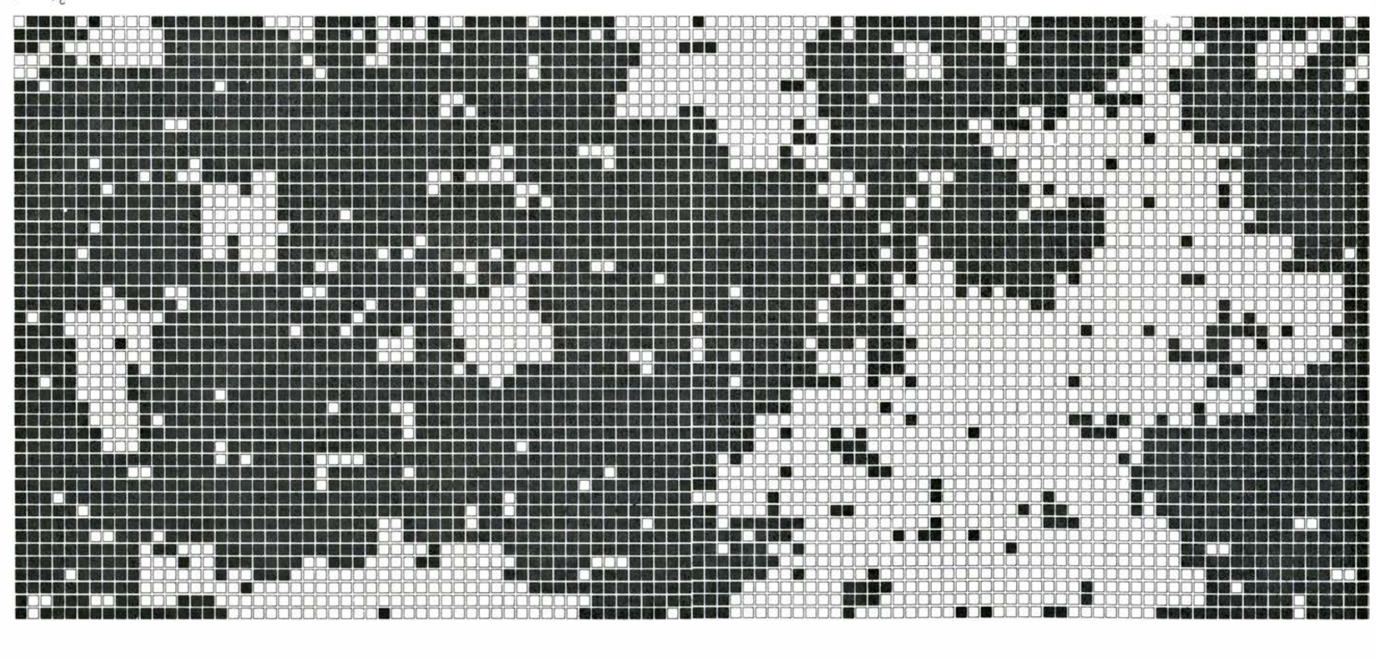
$T = 2T_c$



$T = 1.05T_c$



$T = T_c$



zación es un ferromagneto o imán permanente. Los materiales ferromagnéticos tienen un punto crítico llamado punto de Curie o temperatura de Curie, en honor de Pierre Curie, quien estudió la termodinámica de los ferromagnetos hacia el año 1900. Para el hierro, la temperatura de Curie es de 1044 grados Kelvin. A temperaturas mayores, el hierro no presenta imanación espontánea. Cuando el hierro se enfría, la magnetización se mantiene nula hasta que se alcanza la temperatura de Curie, y entonces el material se magnetiza bruscamente. Si la temperatura se reduce más, la magnetización aumenta suavemente.

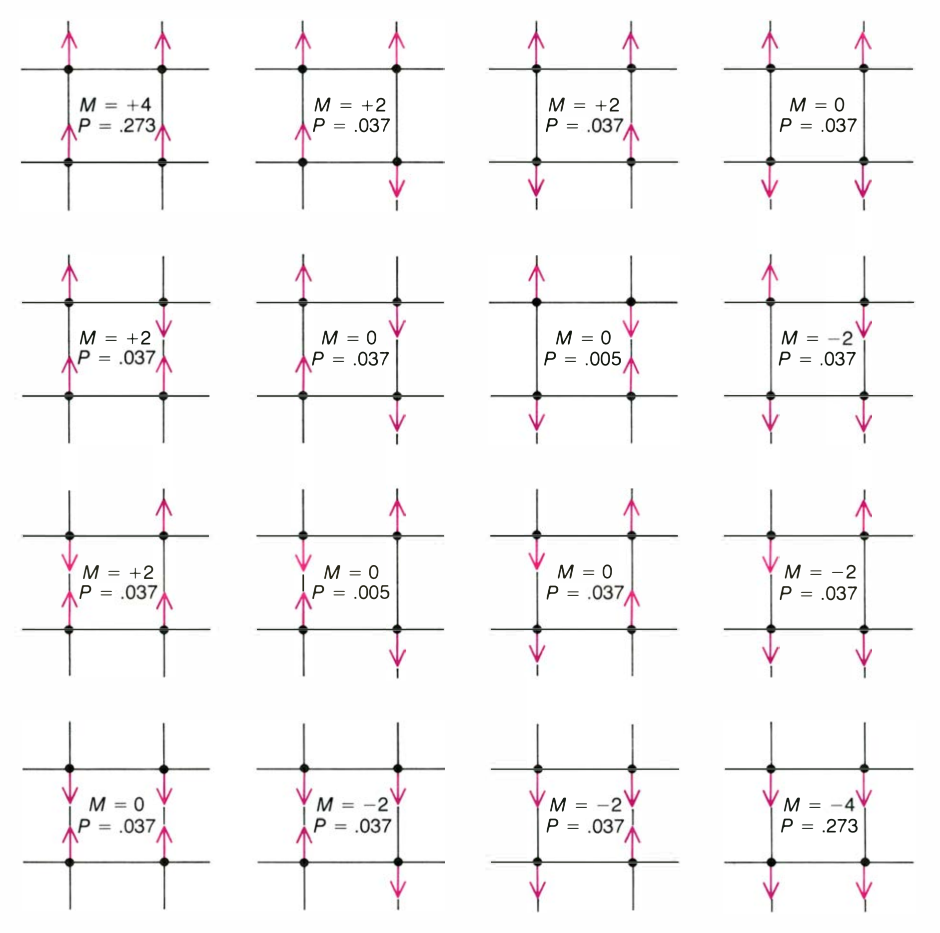
Varias propiedades de los ferromagnetos, aparte de la magnetización, se comportan extrañamente cerca del punto de Curie. Una propiedad de interés es la suceptibilidad magnética, o cambio de la magnetización inducido por la aplicación de un pequeño campo. Bastante por encima del punto de Curie, la susceptibilidad es pequeña porque el hierro no puede mantener ninguna magnetización; bastante por debajo de la temperatura de Curie, la susceptibilidad es pequeña de nuevo porque el material ya está imanado y la aplicación de un campo débil no puede cambiar mucho el estado del sistema. A temperaturas próximas a 1044 grados, sin embargo, la susceptibilidad presenta un pico agudo, y en el mismo punto de Curie la susceptibilidad se hace infinita.

La causa última del ferromagnetismo es el giro mecánico cuántico de los electrones. Debido a que cada electrón gira, tiene un pequeño momento dipolar magnético; en otras palabras, actúa como un imán con un polo norte y un polo sur. No vamos a detallar cómo el espín del electrón da lugar al momento magnético. Baste con advertir que tanto el espín como el momento magnético pueden representarse por un vector, o flecha, que define la dirección del campo magnético del electrón.

Un ferromagneto real tiene una estructura atómica compleja, pero todas las propiedades esenciales del conjunto de espines pueden representarse por un modelo bastante sencillo. En efecto, describiré un modelo que no incluye átomos ni otras partículas materiales sino que consiste sólo en vectores de espín ordenados en una red. Para mayor sencillez, consideraremos una red bidimensional: una cuadrícula rectilínea de líneas en un plano uniformemente espaciadas, con un vector de espín en cada intersección de las líneas de la cuadrícula. Además supondremos que cada espín puede apuntar sólo en dos direcciones, llamadas arriba y abajo. La red del modelo se dice que está imanada cuando más de la mitad de los espines apuntan en la misma dirección. La magnetización puede definirse como el número de espines hacia arriba menos el número de espines hacia abajo.

Cada electrón tiene el mismo espín y el mismo momento dipolar magnético. Lo que distingue a un ferroimán de los demás materiales es un acoplamiento entre espines vecinos que hace que tiendan a alinearse en la misma dirección. Esta tendencia puede establecerse más precisamente indicando que la energía total de cualquier par de espines adyacentes es menor cuando los espines son paralelos que cuando son antiparalelos. La interacción responsable del acoplamiento de los espines es de corto alcance, lo que se refleja en el modelo especificando que sólo los espines vecinos más inmediatos se acoplan entre sí. En la red rectilínea bidimensional cada espín está condicionado por los cuatro vecinos más próximos; ningún otro espín tiene un efecto directo sobre él.

A partir de la naturaleza de la interaccion entre espines en un ferromagneto podría predecirse que todos los espines serían siempre paralelos y el material tendría siempre su máxima imanación. Este es el estado de menor energía, y en ausencia de cualquier efecto perturbador sería el estado más favorable. En un ferroimán real, sin embargo, hay una perturbación que no puede despreciarse: el movimiento térmico de los átomos y de los electrones. A cualquier temperatura por encima del cero absoluto, las excitaciones térmicas del sólido invierten al azar algunos de los espines, y así la dirección del vector de espín cambia, aun cuando la inversión del espín coloque al imán en un estado de mayor energía. Por tanto, no debe

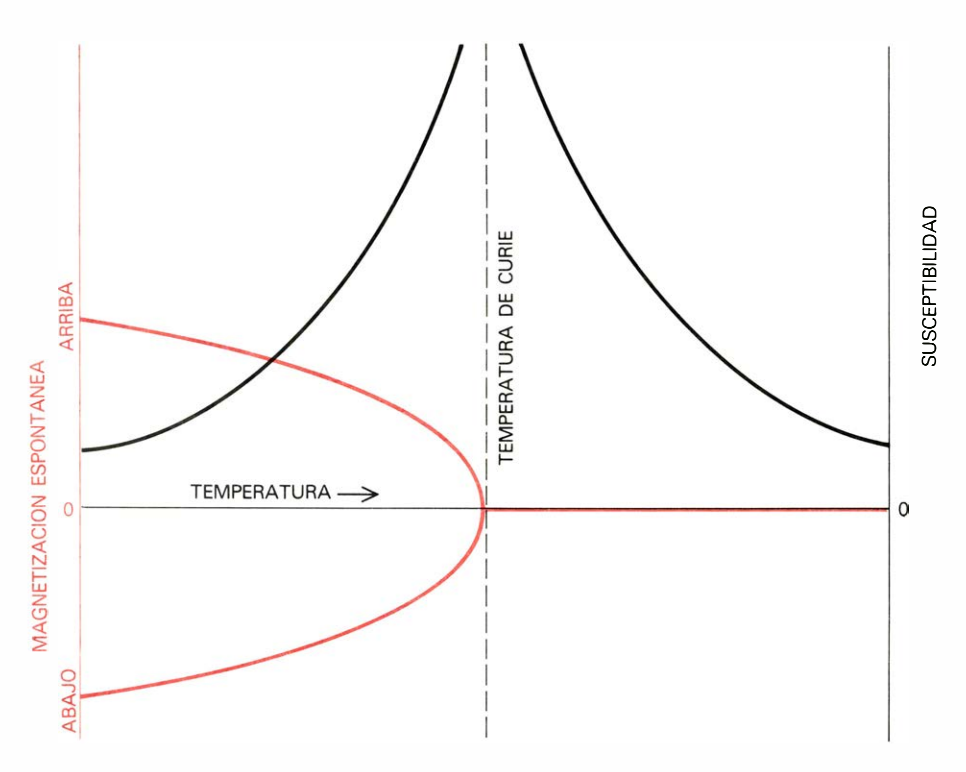


MODELO DE UN FERROIMAN, que consiste en vectores, o flechas de longitud fija, dispuestos en los puntos de una red. Cada vector representa el momento angular de espín y el momento magnético de un solo electrón, y puede orientarse hacia arriba o hacia abajo. Las posiciones vecinas inmediatas de la red están acopladas de tal manera que los vectores de espín adyacentes prefieren ser paralelos, no antiparalelos. A partir de la fuerza de acoplamiento, que decrece cuando la temperatura aumenta, se puede asignar una probabilidad, *P*, a cada configuración posible de los vectores de espín. Todas las configuraciones de una red formada sólo por cuatro puntos se indican aquí. La magnetización resultante, *M*, de cada configuración se calcula fácilmente: es el número de espines hacia arriba menos el número de espines hacia abajo. Para hallar la magnetización a una temperatura dada, se multiplica la magnetización de cada configuración por la probabilidad de dicha configuración, y se suman después los resultados. Las probabilidades se calcularon para una fuerza de acoplamiento de 0,5, que corresponde a una temperatura (en unidades arbitrarias) de 2. El modelo es el bidimensional de Ising.

sorprendernos que la magnetización disminuya cuando aumente la temperatura: dicha relación refleja simplemente el aumento del desorden térmico. Lo que sí resulta curioso es que la imanación no sea una función suave de la temperatura sino que desaparece bruscamente a cierta temperatura, el punto de Curie.

La pugna entre la tendencia hacia una orientación uniforme de los espines y la introducción térmica del desorden puede incorporarse fácilmente en un modelo de ferroimán. La fuerza del acoplamiento entre los espines adyacentes viene dada por un número, K, que debe especificarse al construir el modelo. Los efectos térmicos se incorporan haciendo que K sea inversamente proporcional a la temperatura. Con las unidades de medida apropiadas, la fuerza de acoplamiento puede hacerse igual al inverso de la temperatura, una relación expresada mediante la ecuación $K = 1/T$.

La fuerza del acoplamiento determina la probabilidad de que dos espines adyacentes sean paralelos. Cuando la temperatura es cero, no hay efectos térmicos y los espines adyacentes serán paralelos; la probabilidad es igual a 1 y la fuerza de acoplamiento, infinita. A una temperatura infinita, se anula la fuerza de acoplamiento, de modo que los espines no interaccionan en absoluto. Por tanto, cada espín es libre de escoger su dirección al azar y es independiente de sus vecinos. La probabilidad de que dos espines sean paralelos es 1/2, la misma probabilidad de que sean antiparalelos. La región de interés, desde luego, está entre estos extremos de la temperatura, cuando la probabilidad de que los espines adyacentes estén alineados debe tener siempre un valor entre 1/2 y 1.



MAGNETIZACION de un ferroimán. Se inicia repentinamente a la temperatura de Curie. Por encima de esta temperatura, el número medio de espines hacia arriba y hacia abajo son iguales, y la magnetización se anula. A cualquier temperatura por debajo del punto de Curie, caben dos estados de magnetización, según que predominen los espines hacia arriba o los espines hacia abajo; en ausencia de un campo magnético externo, los dos estados son igualmente probables. La susceptibilidad de un ferroimán o ferromagneto mide el cambio en la magnetización inducido por un campo magnético aplicado, arbitrariamente pequeño. En el punto de Curie, la susceptibilidad se hace infinita. Cerca del punto de Curie, un pequeño cambio en la temperatura o en el campo externo da lugar a otro grande en la magnetización.

Imaginemos una gran red bidimensional de espines en la que un cierto espín se mantenga artificialmente orientado hacia arriba. ¿Qué efecto ejerce sobre los otros espines? El efecto sobre los espines de las cuatro posiciones adyacentes de la red es fácil de imaginar: al hallarse directamente acoplados al espín fijo, tendrán una probabilidad de apuntar hacia arriba mayor que 1/2. Que esa probabilidad no se alcance, y lo que falte para llegar a ella, depende del valor de K, que está determinado a su vez por la temperatura.

Los espines más distantes carecen de interacción directa con el espín fijo; ello no significa que la influencia del espín fijo se acabe en sus vecinos inmediatos. Debido a que los espines más próximos tienden a apuntar hacia arriba con mayor frecuencia que hacia abajo crean una tendencia similar en sus propios vecinos más próximos. De esta manera, la perturbación puede propagarse por gran parte de la red. El alcance de la influencia de un único espín fijo puede medirse observando la orientación de muchos espines alejados y equidistantes del espín fijo. Si la inversión de la orientación del espín fijo de arriba abajo conlleva un aumento del número de espines hacia abajo en la población alejada, entonces se dice que los espines están correlacionados. La distancia máxima a la que una tal correlación puede detectarse se llama longitud de correlación. Las regiones separadas por una distancia mayor que la longitud de correlación son independientes.

En una red a muy alta temperatura, la longitud de correlación es casi nula. La distribución de espines es casi aleatoria, de manera que el número medio de espines arriba y abajo debe ser igual; en otras palabras, la imanación es cero. A medida que la temperatura baja (y la fuerza de acoplamiento aumenta) empiezan a aparecer correlaciones entre distancias mayores. Toman la forma de fluctuaciones de espín, o pequeñas zonas de unos pocos espines que casi todos apuntan en la misma dirección. Contemplada en cualquier área grande, la magnetización sigue siendo nula, pero la estructura de la red es muy distinta de como era cerca de una temperatura infinita.

Cuando la temperatura se acerca al punto de Curie, la longitud de correlación aumenta rápidamente. Las interacciones básicas del modelo no han cambiado; relacionan sólo posiciones adyacentes de la red, pero a partir de las fuerzas de corto alcance se ha establecido un orden de largo alcance. Lo que es más significativo en el aumento de la longitud de correlación es que, a medida que el tamaño máximo de las fluctuaciones de espín aumenta, las fluctuaciones más pequeñas no desaparecen; simplemente se convierten en una estructura más fina superpuesta a una mayor. Las fluctuaciones mayores no son áreas de alineaciones de espines uniformes; comprenden muchas fluctuaciones más pequeñas y pueden distinguirse tan sólo porque tienen un exceso global de una dirección de espín. Así, un océano de espines que están casi todos hacia arriba puede tener una isla de espines que están casi todos hacia abajo y que a su vez rodee un lago de espines hacia arriba con un islote de espines hacia abajo. La progresión continúa hasta la escala más pequeña posible: un solo espín.

Cuando la temperatura alcanza el valor de la de Curie, la longitud de correlación se hace infinita. Cualquier

par de espines están correlacionados, independientemente de cuál sea la distancia entre ellos. Sin embargo, persisten fluctuaciones a todas las escalas de longitud menores. El sistema permanece sin magnetizar, pero es extremadamente sensible a pequeñas perturbaciones. Por ejemplo, mantener un solo espín fijo con dirección hacia arriba provoca una perturbación que se extiende por toda la red y da a todo el sistema una cierta magnetización.

Por debajo de la temperatura de Curie, el sistema se magnetiza incluso en ausencia de una perturbación externa, pero no hay un cambio inmediato en el aspecto de la red. Las fluctuaciones de menor escala persisten; son remanentes de los lagos e islotes de dirección de espín opuesta. Mirando simplemente a la red no puede detectarse la magnetización. Sólo cuando el sistema se enfría más se hace obvia la orientación preferente, ya que el aumento de la fuerza de acoplamiento obliga a más espines a estar como la mayoría. A temperatura cero se alcanza la uniformidad completa.

En los fluidos, las fluctuaciones de la densidad cerca del punto crítico guardan una estrecha analogía con las fluctuaciones en las direcciones de los espines observadas en los ferromagnetos. En los fluidos, sin embargo, puede observarse directamente la presencia de fluctuaciones a todas las posibles escalas de longitud. Cuando la longitud de correlación llega a alcanzar unos pocos miles de angstrom, lo que es comparable a la longitud de onda de la luz, las fluctuaciones empiezan a dispersar la luz fuertemente y el fluido se hace lechoso, fenómeno que se conoce por opalescencia crítica. De un modo significativo, cuando la temperatura se aproxima aún más al punto crítico y la máxima escala de fluctuaciones se hace mucho mayor (milímetros o centímetros), no se reduce la opalescencia crítica, lo que indica que persisten las fluctuaciones menores. El mismo fenómeno tiene lugar en sistemas de espines, pero debido a que los materiales ferromagnéticos no son transparentes a la luz, no puede mostrarse tan fácilmente. La opalescencia crítica de los ferroimanes se ha detectado, no obstante, en la dispersión de neutrones por un material magnético cerca de la temperatura de Curie.

El modelo que he descrito no es una invención mia. Es una versión del que introdujeron en los años 20 los físicos alemanes Wilhelm Lenz y Ernest Ising, denominado modelo de Ising. Las propiedades de un sistema de espines de Ising en una red bidimensional se conocen en todo detalle porque el modelo fue resuelto de un modo exacto en 1944 por Lars Onsager, de la Universidad de Yale. Desde entonces se han encontrado también soluciones para otros varios modelos bidimensionales. (Todavía no se ha resuelto con exactitud ningún modelo tridimensional.) Sin embargo, los problemas de describir sistemas bidimensionales están lejos de ser triviales. En lo que sigue aplicaré los métodos del grupo de renormalización al modelo de Ising bidimensional como si fuese un problema aún no resuelto; la solución de Onsager servirá como comprobante de los resultados.

¿Qué significa resolver o entender un modelo de un sistema físico? En el caso del sistema de Ising, las propiedades microscópicas se conocen en toda su extensión desde el inicio, pues se especificaron al construir el modelo. Se necesita un medio de predecir las propiedades macroscópicas del sistema a partir de las microscópicas. Por ejemplo, en una fórmula que dé la magnetización espontánea, la susceptibilidad y la longitud de correlación del modelo en función de la temperatura contribuiría mucho a la comprensión.

No entraña una dificultad particular calcular las propiedades macroscópicas de una configuración dada de los espines en un modelo de Ising. La magnetización, por ejemplo, puede determinarse con sólo contar el número de espines hacia arriba y el número de espines hacia abajo y restándolos. Pero ninguna configuración de los espines determina las propiedades macroscópicas del sistema. Por el contrario, todas las posibles configuraciones contribuyen a las propiedades observadas, cada una proporcionalmente a su probabilidad a una temperatura dada.

En principio, las propiedades macroscópicas podrían calcularse directamente como la suma de todas las contribuciones por separado. Primero se hallaría la magnetización para cada configuración y luego su probabilidad correspondiente. La magnetización real se encontraría multiplicando cada uno de estos pares de números y sumando todos los resultados. La susceptibilidad y la longitud de correlación podrían deducirse por procedimientos que no son mucho más elaborados. El elemento común en todos estos cálculos es la necesidad de determinar las probabilidades de todas las configuraciones posibles de los espines. Una vez se conoce la distribución de probabilidades, las propiedades macroscópicas se deducen directamente.

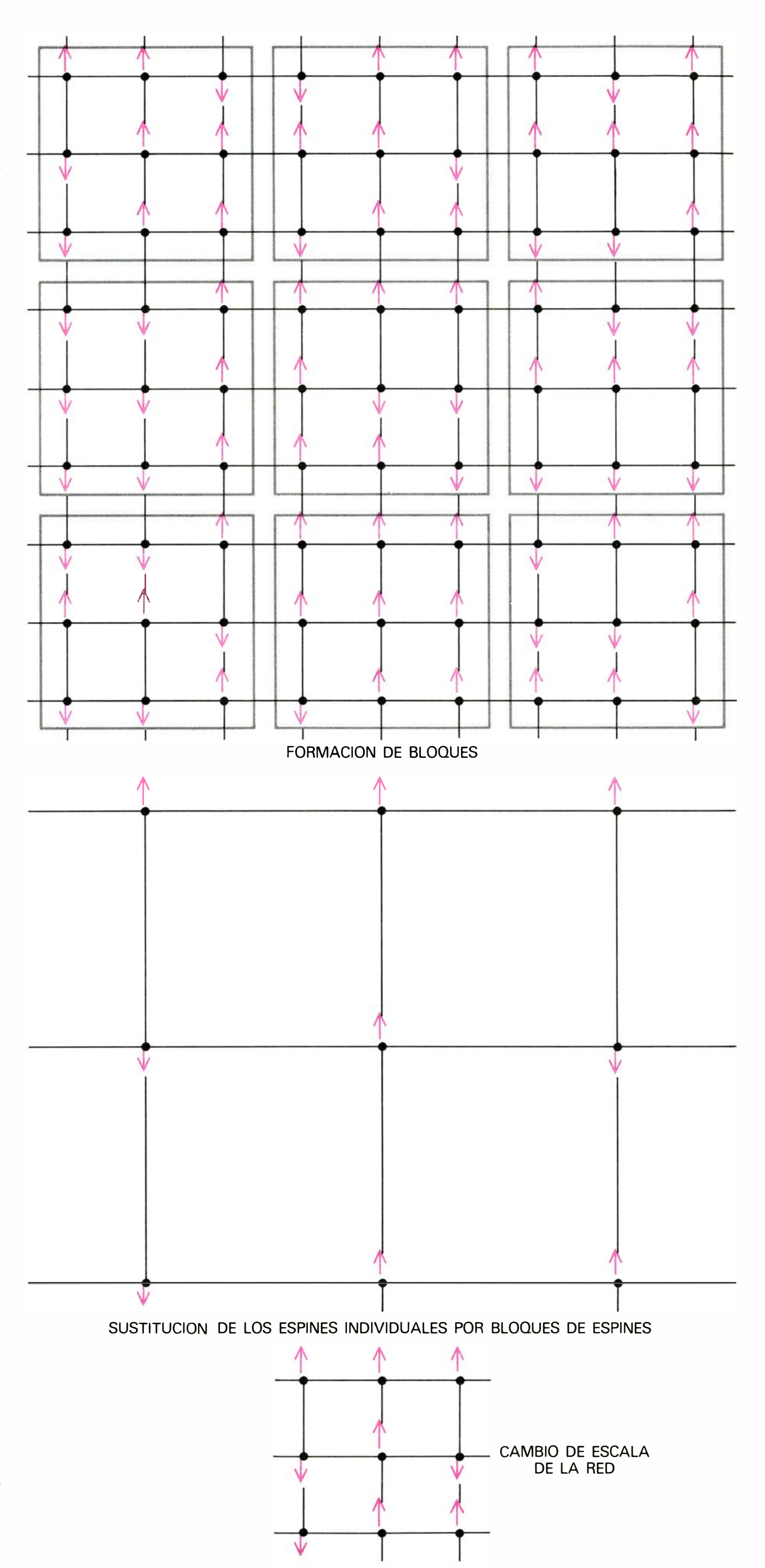
NUMERO DE CONFIGURACIONES DE ESPINES / TAMAÑO DE LA RED
1 × 10^30, 2 × 10^24, 2 × 10^19, 6 × 10^14, 7 × 10^10, 3 × 10^7, 65,536, 512, 16, 2
1 × 1, 2 × 2, 3 × 3, 4 × 4, 5 × 5, 6 × 6, 7 × 7, 8 × 8, 9 × 9, 10 × 10

NUMERO DE CONFIGURACIONES DE ESPINES: aumenta rápidamente cuando el tamaño de la red crece. Para un sistema de *n* espines, en que cada uno pueda tomar dos valores posibles, el número de configuraciones es igual a 2^n. Cuando la red es grande, no se puede calcular la probabilidad de todas las configuraciones. El límite de cálculo práctico está en una red algo mayor que el conjunto seis por seis de 36 espines. A fin de observar el comportamiento crítico del sistema cerca de la temperatura de Curie, se necesitaría un conjunto de unos 100 por 100 espines, que tiene $2^{10.000}$ configuraciones.

Como he indicado antes, la probabilidad de que cualquier par de espines adyacentes sean paralelos se determina únicamente por la fuerza de acoplamiento K, definida como el recíproco de la temperatura. Si llamamos p a la probabilidad de que dos espines vecinos aislados sean paralelos, la probabilidad de que sean antiparalelos debe ser $1 - p$. A partir de estos dos valores puede evaluarse la probabilidad relativa de cualquier configuración dada de una red. Todo lo que se precisa es multiplicar las probabilidades separadas para cada par de espines vecinos más próximos, tomando en cada caso el valor p si los espines son paralelos y $1 - p$ si son antiparalelos.

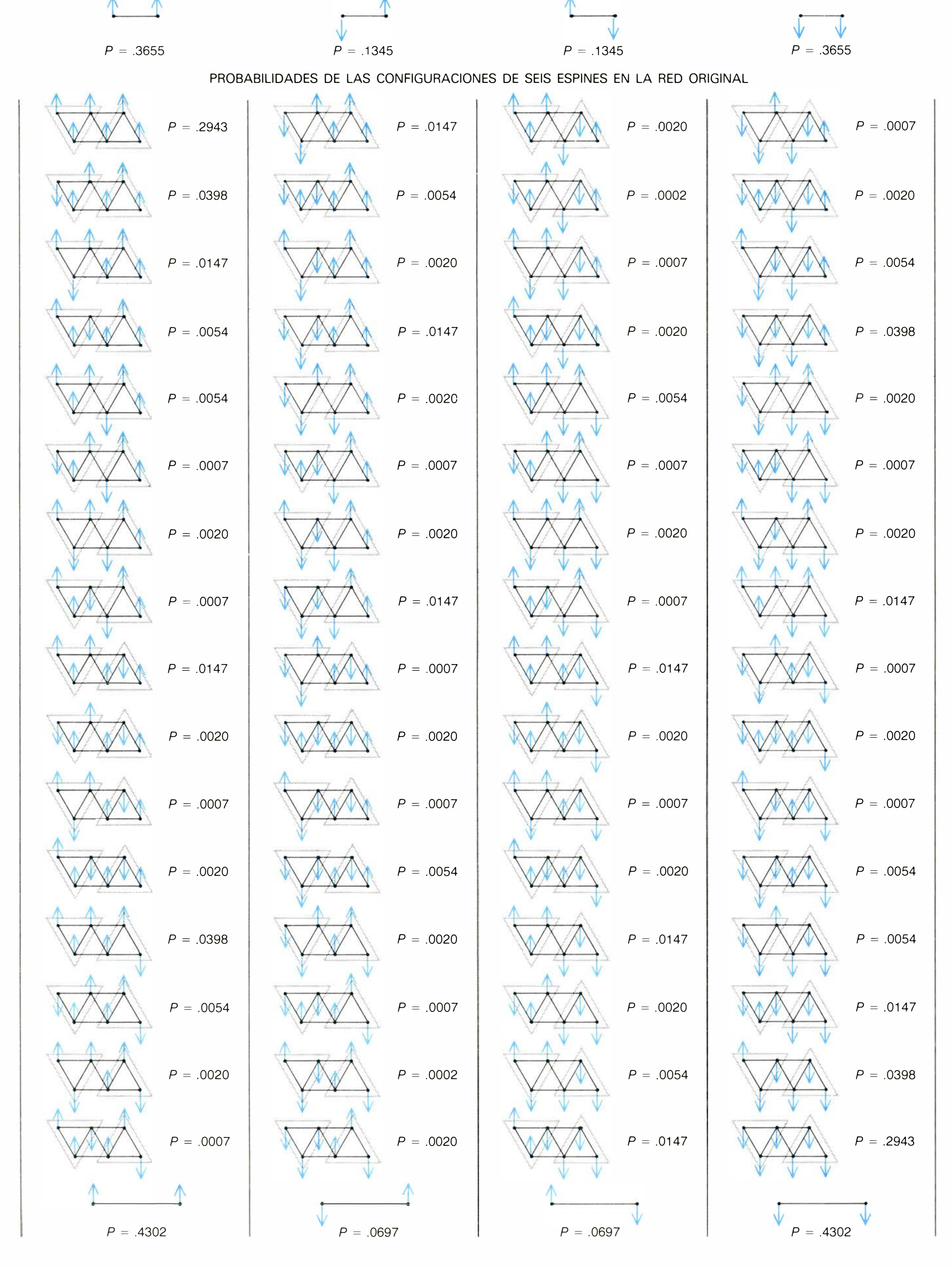
Consideremos un sistema de espines que esté formado por cuatro de ellos colocados en los ángulos de un cuadrado. Una tal red tiene cuatro acoplamientos entre vecinos más próximos, correspondientes a los cuatro lados del cuadrado. Cada acoplamiento se considera por separado y se le asigna una probabilidad p o $1 - p$, según los espines sean paralelos o antiparalelos; se multiplican luego las cuatro probabilidades. En la configuración con los cuatro espines orientados hacia arriba, los cuatro pares son paralelos y la probabilidad relativa viene dada por el producto $p \times p \times p \times p$. Si tres espines están hacia arriba y uno hacia abajo, la probabilidad relativa es $p \times p \times (1 - p) \times (1 - p)$.

El cálculo debe hacerse para cada configuración de los espines; para un sistema de cuatro espines hay 16 configuraciones. Un último paso es convertir las probabilidades relativas en absolutas ajustando cada valor de manera que el total de los 16 valores sea exactamente 1. Como la temperatura determina la fuerza de acoplamiento y la fuerza de



APROXIMACION del grupo de renormalización a un modelo de ferroimán. Consiste en dividir un problema intratable con múltiples escalas de longitud en una serie de problemas más pequeños, cada uno de los cuales se limita a una única escala de longitud. Una versión del método del grupo de renormalización, llamada transformación de bloques de espines, tiene tres pasos. Primero la red se divide en bloques de unos pocos espines cada uno, en este caso nueve espines. Después, cada bloque se sustituye por un único espín cuyo valor es el promedio de todos los espines del bloque; aquí el promedio se determina por la regla de la mayoría. De esta manera, se crea una nueva red con un espaciado tres veces mayor que el de la red original y una densidad de espines de un tercio de la densidad de partida. Finalmente, se restituye la escala original reduciendo todas las dimensiones por un factor 3. El proceso debe hacerse para todas las configuraciones de unos pocos espines de la red original, de manera que podamos calcular una probabilidad para cada configuración de los espines del bloque.

PROBABILIDADES DE LAS CONFIGURACIONES DE VECINOS INMEDIATOS EN LA RED ORIGINAL



PROBABILIDADES DE LAS CONFIGURACIONES DE BLOQUES DE ESPINES VECINOS INMEDIATOS

acoplamiento, a su vez, determina los valores de p y de $1 - p$, la serie completa de los 16 cálculos debería repetirse para cada temperatura de interés.

Este plan de ataque del modelo de Ising es ambicioso pero no es práctico. Si la probabilidad de cada configuración de espines pudiera calcularse, la magnetización y las otras propiedades macroscópicas podrían evaluarse para cada temperatura dada. El problema reside en el número de configuraciones de espín. Para un sistema formado por n espines, cada uno de los cuales puede tomar uno de dos valores hay 2^n configuraciones posibles. Esta función exponencial aumenta rápidamente cuando n crece. Como hemos indicado, cuatro espines tienen 2^4, o 16, configuraciones. Un bloque tres por tres de nueve espines tiene 512 configuraciones y un bloque cuatro por cuatro tiene 65.536. El límite práctico de cálculo no es mucho mayor que un bloque seis por seis de 36 espines, para el cual hay aproximadamente 7×10^{10} configuraciones.

¿Qué tamaño debe tener la red a fin de determinar las propiedades críticas del modelo de Ising bidimensional? La red debe ser por lo menos tan grande como las mayores fluctuaciones observadas a la temperatura de interés. A una temperatura razonablemente próxima al punto de Curie, la longitud de correlación en unidades del espaciado de la red puede ser de unos 100 y las mayores fluctuaciones cubrirían unos 100^2, o 10.000, puntos de la red. Un bloque de espines de este tamaño tiene $2^{10.000}$ configuraciones posibles, un número que es algo mayor que 10^{3000}. El computador más rápido concebible no podría realizar un cálculo semejante. Aun cuando el computador hubiera estado trabajando sin cesar desde la "gran explosión" ("big bang") en que se inició el universo, no habría hecho un avance significativo en el cálculo.

La necesidad de realizar una enumeración casi sin fin de configuraciones de espines puede evitarse para dos condiciones especiales de la red. Cuando la temperatura del sistema es cero (de manera que la fuerza de acoplamiento es infinita) pueden despreciarse todas las configuraciones menos dos. A temperatura nula, la probabilidad de que un par de espines sean antiparalelos es cero, y por tanto también se anula la probabilidad de cualquier configuración que incluya algún par antiparalelo. Las únicas configuraciones que no tienen al menos un par antiparalelo son aquellas en que todos los espines están hacia arriba o todos hacia abajo. La red tomará con seguridad una de estas configuraciones, y todas las otras configuraciones poseerán probabilidad cero.

A temperatura infinita, cuando la fuerza de acoplamiento es cero, la distribución de probabilidad también se simplifica mucho. Cada espín es independiente de sus vecinos y su dirección en cualquier instante puede escogerse al azar. El resultado es que cada configuración de la red tiene la misma probabilidad.

Con estas dos sencillas maneras de determinar la distribución de probabilidad, constituye un ejercicio trivial calcular exactamente las propiedades del modelo de Ising en el cero absoluto y a temperatura infinita. También hay métodos aproximados aceptables para cualquier temperatura lo suficientemente baja para ser considerada cercana al cero o suficientemente alta para considerarla cercana al infinito. La región problemática se sitúa entre estos extremos; corresponde a la región del punto crítico. Hasta hace poco no había un método práctico y directo de calcular las propiedades de un sistema arbitrariamente cercano al punto crítico. El grupo de renormalización proporciona un tal método.

La esencia del método del grupo de renormalización consiste en desglosar un problema sumamente complejo en una serie de pasos menores y más manejables. En vez de considerar todos los espines en una región del tamaño de la longitud de correlación, las propiedades de largo alcance se deducen a partir del comportamiento de unas pocas cantidades que incorporan los efectos de muchos espines. Hay varias maneras de hacer esto. Describiré una, la técnica de los bloques de espines, en la que los principios del método aparecen particularmente claros. Introducida por Leo P. Kadanoff de la Universidad de Chicago, fue convertida en una herramienta práctica para los cálculos por Th. Niemeijer y J. M. J. van Leeuwen, de la Universidad de Tecnología de Delft, Holanda.

El método consta de tres pasos fundamentales, cada uno de los cuales debe repetirse muchas veces. Primero la red se divide en bloques con unos pocos espines en cada uno. Emplearemos bloques cuadrados de tres espines de lado, de manera que cada bloque contenga nueve espines. Después, todos los espines del bloque se promedian de alguna manera y se sustituye todo el bloque por un único nuevo espín con el valor promedio. En este paso, el promedio puede hacerse de una manera sencilla: siguiendo el principio de la mayoría. Si cinco o más de los espines originales están hacia arriba, el nuevo espín está también hacia arriba; de lo contrario, está hacia abajo.

El resultado de estas dos operaciones ha sido la creación de una nueva red cuyo espaciado fundamental es tres veces mayor que el de la red inicial. En el tercer paso se restablece la escala de partida reduciendo todas las dimensiones en un factor 3.

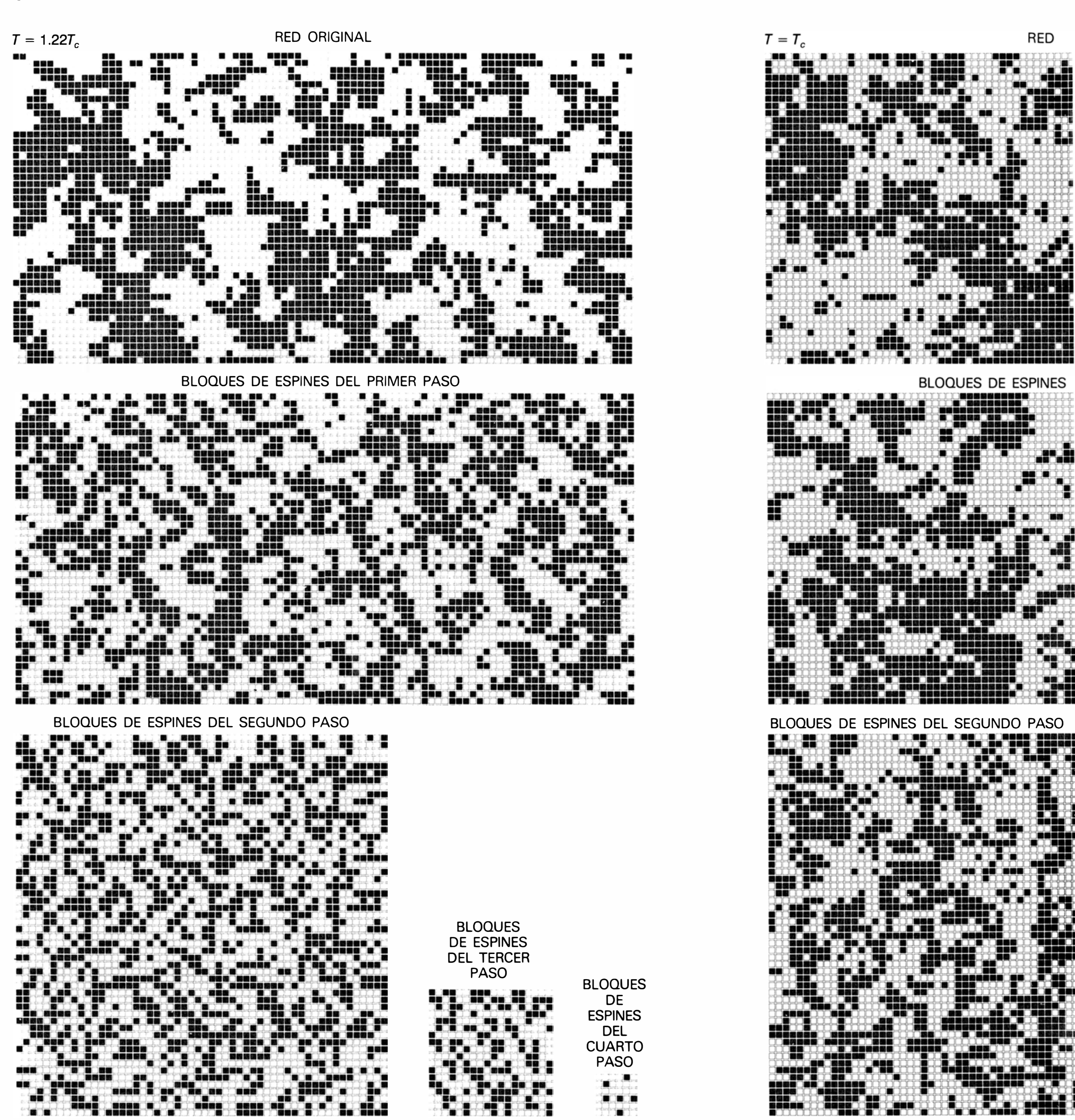
Estos tres pasos definen una transformación del grupo de renormalización. Su efecto es eliminar del sistema todas las fluctuaciones de las direcciones de los espines cuya escala es más pequeña que el tamaño del bloque. En el modelo presentado, cualquier fluctuación de los espines de un alcance igual o menor que tres unidades de la red será suavizada por el promedio de los espines en cada bloque. Es como si mirásemos la red a través de una lente desenfocada, en que los detalles pequeños se debilitan pero los mayores no quedan afectados.

No basta hacer esta operación para cualquier configuración de la red original; una vez más, lo que se busca es una distribución de probabilidad. Supongamos que se considera sólo una pequeña región de la red inicial, que conste de 36 espines que puedan distribuirse en cuatro bloques. Los espines de esta región tienen 2^{36} configuraciones posibles, es decir, unos 70.000 millones. Realizada la transformación en bloques de espín, los 36 espines originales quedan sustituidos por cuatro bloques de ellos con un

DISTRIBUCION DE PROBABILIDAD para un sistema de bloques de espines. Se halla sumando las probabilidades de todas las configuraciones de la red original que contribuyen a cada configuración de los bloques de espines. Se muestra el cálculo para un sistema de seis espines en una red triangular. A partir de la red, se forman dos bloques de tres espines cada uno y cada bloque se sustituye por un único espín cuya orientación se determina por la regla de la mayoría. Los seis espines tienen 64 posibles configuraciones, que se asignan a una columna de tal manera que todas las configuraciones de cada columna dan lugar a la misma configuración de bloques de espines. Por ejemplo, todas las configuraciones de la columna de la izquierda tienen al menos dos espines en cada bloque apuntando hacia arriba y así se representan por dos bloques de espín hacia arriba. La fuerza de acoplamiento en la red original se toma igual a 0,5, lo que da las probabilidades entre vecinos inmediatos mostradas en la parte de arriba de la página precedente. A partir de este conjunto de números se calcula una probabilidad para cada configuración de la red original; después se suman todas las probabilidades en cada columna para obtener la probabilidad de la configuración de bloques de espines correspondiente. Las probabilidades de los bloques de espines no son las mismas que las especificadas para la red original, lo que implica que la fuerza de acoplamiento sea también distinta, como difiere también la misma temperatura.

total de 16 configuraciones. El calcular la probabilidad de cada una de las configuraciones de los 36 espines originales está justo dentro de los límites de lo factible. A partir de estos números pueden determinarse fácilmente las probabilidades de las configuraciones de los 16 bloques de espines. El cálculo puede hacerse clasificando todas las configuraciones de la red original en 16 grupos según qué configuración de bloques de espines resulte en cada caso cuando se aplique la regla de la mayoría. La probabilidad total de una cierta configuración de bloques de espín se halla sumando las probabilidades de todas las configuraciones de la red original que pertenecen a tal grupo.

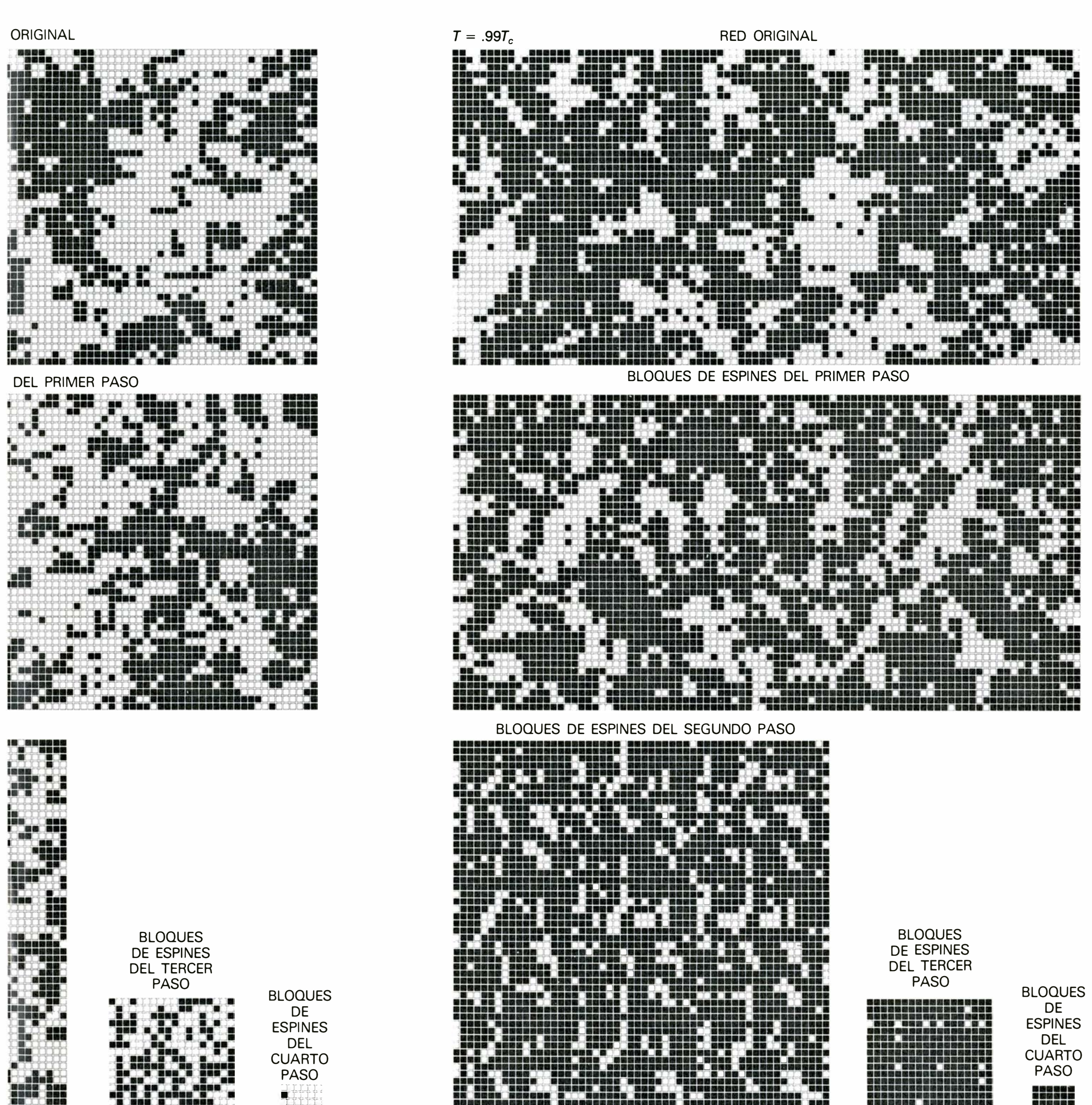
Pudiera parecer que no se gana nada con este procedimiento. Si puede calcularse la distribución de probabilidades completa para un sistema de 36 espines, no se aprende nada nuevo condensando el sistema en una red más pequeña de



TRANSFORMACION de bloques de espines. Se aplica repetidamente a una red de bloques de espines, elucidando cada vez el comportamiento del sistema a una escala mayor. La simulación del computador empezó con un conjunto de unos 236.000 espines; un cuadrado negro representa un espín hacia arriba y un cuadrado blanco un espín hacia abajo. La temperatura inicial se tomó igual a tres valores: uno por encima de la temperatura de Curie, T_c, otro a T_c, y otro justo debajo de T_c. La transformación empieza con la división de la red en bloques tres por tres. Cada bloque se sustituye por un único espín cuyo valor se determina por la regla de la mayoría; éstos constituyen la red de bloques de espines de la primera etapa. El método se repite luego, pero tomando como red inicial los bloques de espines de la primera etapa. Los espines de la segunda etapa resultantes forman la configuración original para la siguiente transformación, y así sucesivamente. Cuando se alcanza la tercera etapa, el número de los espines es lo suficien-

cuatro bloques de espín. Cerca del punto crítico aún es necesario considerar una red mucho mayor, quizá con 10.000 espines en vez de 36, y la distribución de probabilidad para los bloques de espín obtenidos a partir de esta red no puede calcularse porque hay aún muchísimas configuraciones. Sin embargo, resulta que hay un método para extraer información útil a partir de un pequeño conjunto de bloques de espines. Se trata de un método para observar el comportamiento del sistema en una zona grande sin necesidad de tratar explícitamente las configuraciones de todos los espines de la zona.

Cada bloque de espines representa nueve espines de la red original. Sin embargo, el conjunto completo de los bloques puede contemplarse también como un verdadero sistema de espines, cuyas propiedades pueden estudiarse con los mismos métodos que se aplican al modelo original. Puede suponerse que haya



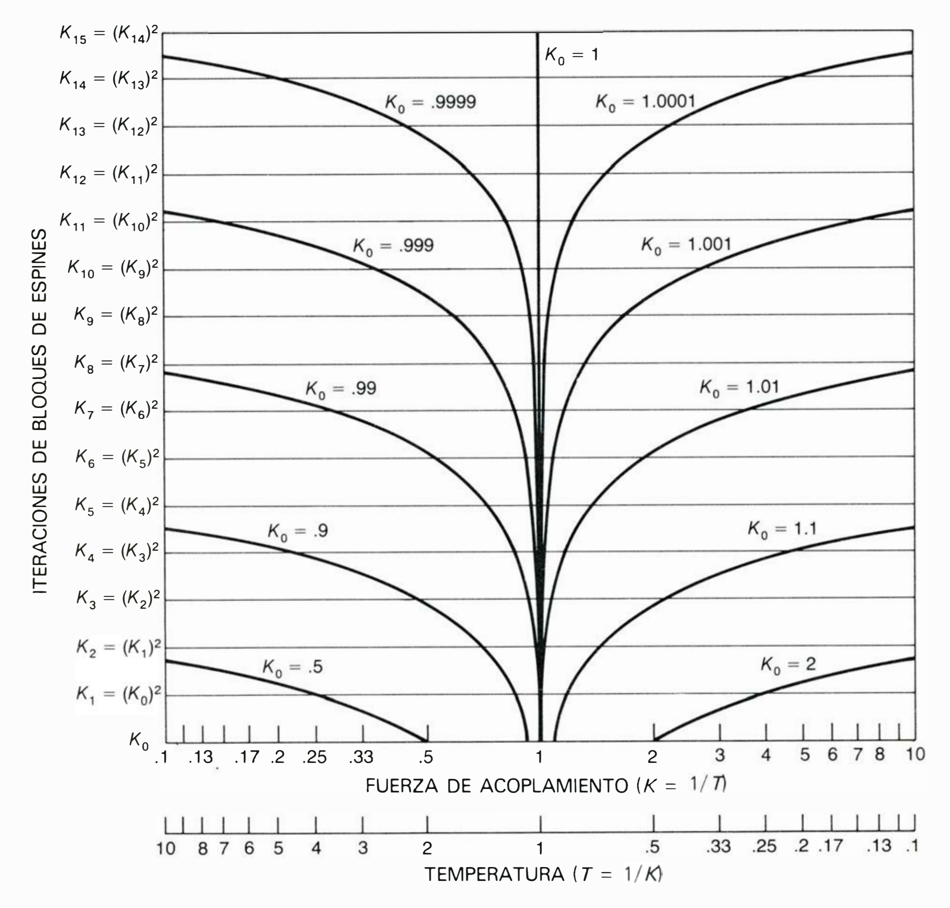
temente pequeño para que puedan representarse todos; después de la cuarta etapa sólo quedan 36 espines, cada uno de los cuales representa más de 6000 posiciones de la red original. En la primera etapa, cualquier fluctuación cuya escala de longitud sea menor que tres unidades de la red se elimina al promediar. La segunda etapa elimina las fluctuaciones de entre tres y nueve unidades de la red, la tercera etapa las situadas entre nueve y 27 unidades, etcétera. Cuando la temperatura inicial es superior a T_c los espines aparentan distribuirse más al azar con cada iteración, y desaparecen las fluctuaciones de gran escala; cuando la está por debajo de T_c, los espines se distribuyen más uniformemente y las fluctuaciones que permanecen son de menor escala. Si la temperatura inicial es exactamente igual a T_c, las fluctuaciones de gran escala permanecen en todas las etapas. Debido a que la transformación de bloques de espines no cambia la estructura de gran escala de la red a la temperatura de Curie, está en un punto fijo.

acoplamientos entre los bloques de espines, que dependen de la temperatura y que determinan a su vez la probabilidad de cada posible configuración de espín. Una hipótesis de partida puede ser que los acoplamientos entre los bloques de espín son los mismos que se especificaron en la red original de espines de Ising, es decir, una interacción entre vecinos inmediatos con una fuerza dada por el parámetro K, el inverso de la temperatura.

Esta hipótesis puede comprobarse fácilmente porque ya se conoce la distribución de probabilidades de las configuraciones de al menos una pequeña parte del sistema de bloques de espín; se calculó a partir de las configuraciones de la red original cuando se definieron los bloques de espines. Sorprendentemente, esta hipótesis en general es incorrecta: los bloques de espines no tienen los mismos acoplamientos que los espines del modelo original. Suponer que sólo las posiciones adyacentes interaccionan y que tienen la misma fuerza de acoplamiento K, proporciona un conjunto de probabilidades equivocado para las configuraciones de los bloques de espines.

Si las especificaciones del modelo original no describen el sistema de bloques de espines, debe inventarse algún nuevo conjunto de acoplamientos. El principio fundamental para formular estas nuevas interacciones consiste en reproducir tan aproximadamente como sea posible las distribuciones de probabilidad observadas. En general, la fuerza de acoplamiento entre vecinos inmediatos debe cambiarse, es decir, K debe tomar un nuevo valor. Aún más, deben introducirse acoplamientos de más largo alcance, que se excluyeron por definición en el modelo de Ising. Por ejemplo, puede ser necesario establecer un acoplamiento entre los espines de los ángulos opuestos de un cuadrado. Pueden haber también interacciones directas entre grupos de tres o cuatro espines. Son posibles acoplamientos de alcance todavía mayor. Por tanto, los bloques de espines pueden considerarse como un sistema reticular, aunque se trata de un sistema bastante distinto del original. Especialmente, debido a que los acoplamientos básicos tienen distintos valores, la red de bloques de espines está a una temperatura distinta de la del sistema de Ising inicial.

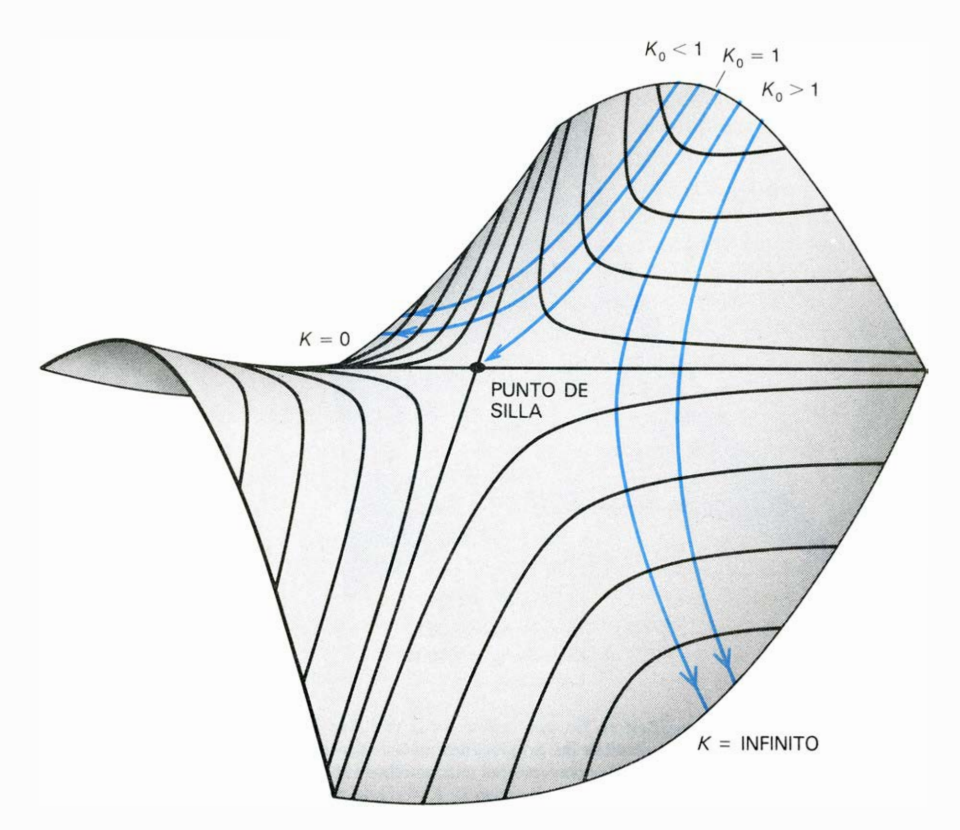
Una vez se ha encontrado un conjunto de acoplamientos que describa correctamente la distribución de probabilidad de los bloques de espines, puede construirse con ellos una red de tamaño arbitrario. La nueva red se forma del mismo modo que la original, pero ahora la probabilidad para el espín de cada posición está determinada por las fuerzas de acoplamiento recién obtenidas en vez de estarlo por el único acoplamiento del modelo de Ising. El cálculo del grupo de renormalización continúa ahora, empezando otra vez el proceso a partir del nuevo sistema de bloques de espines. Igual que en el caso anterior, se forman bloques de nueve espines cada uno y en una región pequeña, un conjunto de 36 espines, por ejemplo, se halla la probabilidad de cada configuración posible. Este cálculo se usa luego para definir la distribución de probabilidad de una segunda generación de bloques de espines que se forman de nuevo por la regla de la mayoría. El análisis de los bloques de espines de la segunda generación muestra que los acoplamientos han cambiado de nuevo, de manera que deben suministrarse por segunda vez nuevos valores para cada fuerza de acoplamiento. Una vez se han determinado estos nuevos valores puede construirse otro sistema reticular (la tercera generación) y todo el procedimiento puede repetirse de nuevo.



CAMBIO EN EL ACOPLAMIENTO entre los espines. Dicho cambio es parte de la transformación del grupo de renormalización. El ajuste que debe hacerse a la fuerza de acoplamiento en cada iteración puede tomar muchas formas, pero aquí se presenta un ejemplo sencillo: Si la fuerza de acoplamiento en la red original se da por un número K, en la nueva red la fuerza de acoplamiento es igual a K^2. Cualquier valor inicial de K mayor que la unidad debe tender a infinito cuando repetidamente se eleva al cuadrado; cualquier valor menor que la unidad debe tender a cero. El valor especial $K = 1$ permanece invariable, independientemente de cuantas veces se repita la transformación. Puesto que la temperatura puede definirse (en unidades apropiadas) como el inverso de la fuerza de acoplamiento, la transformación del grupo de renormalización puede contemplarse como si se estableciera una correspondencia entre la red original y la nueva red más delgada que, en general, tendrá una fuerza de acoplamiento distinta y una temperatura diferente. Sólo en el punto fijo, punto que corresponde a la temperatura de Curie, el acoplamiento y la temperatura permanecen invariables con valor unidad.

Lo importante de esta operación repetitiva es que proporciona información acerca del comportamiento de sistemas de espines distintos, aunque relacionados, en los que la escala fundamental de longitud crece con cada iteración. Después de la primera transformación de bloques de espín se han eliminado las fluctuaciones de la menor escala, pero pueden verse más claramente las ligeramente mayores, correspondientes a una escala de unas tres veces el espaciado de la red original. Después de la segunda transformación, cada bloque espín representa los 81 espines de un bloque nueve por nueve de la red original, y todas las fluctuaciones de alcance menor que este tamaño se han promediado, dejando sólo las mayores que nueve unidades de la red. La siguiente iteración

elimina todas las fluctuaciones cuya escala está entre nueve y 27 unidades de la red; la siguiente iteración elimina después las que están entre 27 y 81 unidades. A la larga quedan promediadas las fluctuaciones de todas las escalas hasta la longitud de correlación. El sistema de espines resultante refleja sólo las propiedades de largo alcance del sistema de Ising original, con todas las fluctuaciones de escala menor eliminadas.



EVOLUCION DE UN SISTEMA DE ESPINES en respuesta a repetidas transformaciones del grupo de renormalización; puede describirse como el movimiento de un punto en una superficie construida en un espacio imaginario multidimensional: el espacio de los parámetros. La forma de la superficie está definida por todos los acoplamientos entre bloques de espines, pero aquí sólo se considera el acoplamiento vecino inmediato, *K*. La superficie tiene dos picos y dos simas que se empalman en un punto de silla. La trayectoria seguida por el punto que representa el estado del sistema está determinada completamente por la pendiente de la superficie. Un valor inicial de *K* ligeramente mayor que 1 corresponde a una posición inicial justo a un lado de la cresta que conecta los picos. Tras varias transformaciones de bloques de espín, el punto cae por la colina, pasa cerca del punto de silla y se aleja hacia una de las simas, donde el valor de *K* tiende hacia infinito. Un valor inicial de *K* ligeramente menor que la unidad conduce a una trayectoria similar por el otro lado de la cresta y acaba en la otra sima donde *K* tiende a cero. Cuando el valor de *K* es exactamente la unidad, el punto se instala permanentemente sobre la cresta, acercándose al equilibrio en el punto de silla. Las dos simas son puntos fijos (ya que los valores de $k = 0$ y $K =$ infinito no cambian con transformaciones adicionales del grupo de renormalización), pero se consideran puntos fijos triviales. El punto de silla define el punto fijo crítico (Gabor Kiss.)

La potencia de la técnica de bloques de espines puede verse incluso mediante una simple ojeada sobre la evolución del modelo. Si nos fijamos en una configuración de espines de Ising justo debajo de la temperatura de Curie, difícilmente se verá que el modelo está ligeramente imanado. A esta temperatura hay sólo un pequeño exceso de una dirección de espín sobre la otra y las numerosas fluctuaciones de pequeña escala obscurecen la tendencia general. Después de varias aplicaciones de la transformación de bloques de espines, sin embargo, las fluctuaciones más pequeñas desaparecen y se ve claramente la magnetización de largo alcance.

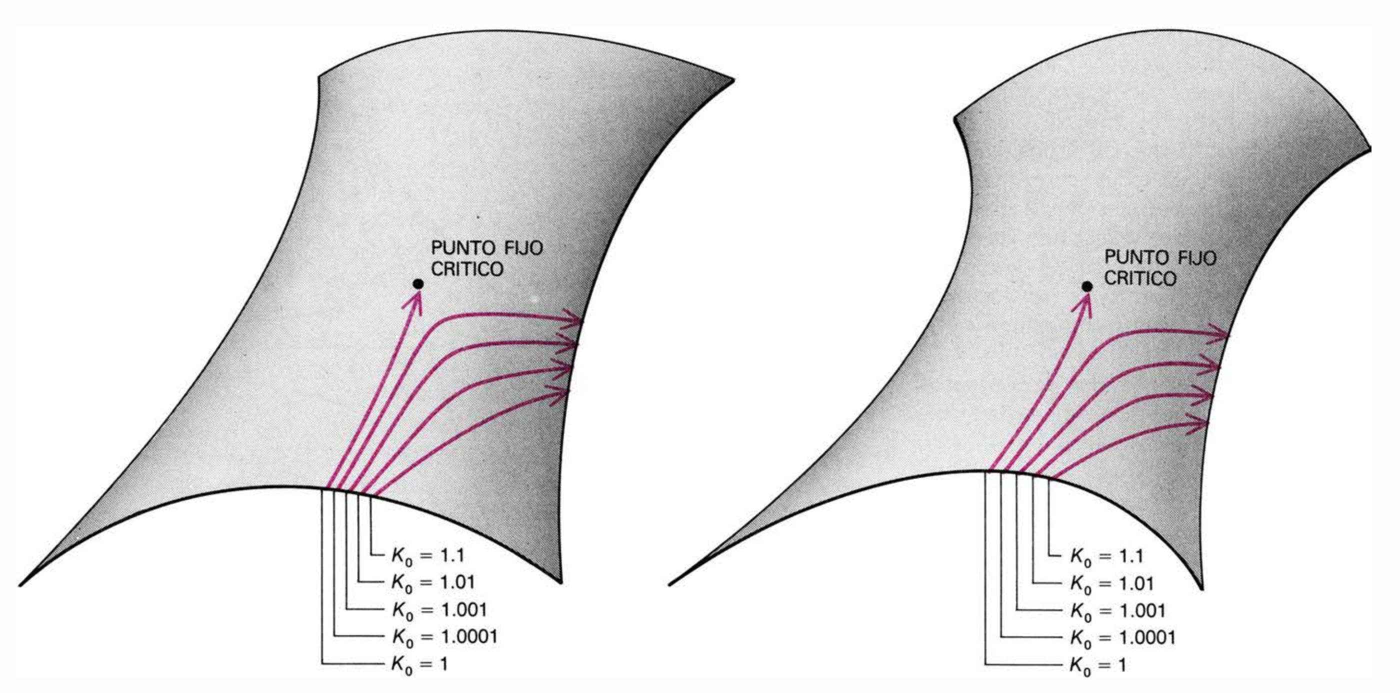
Una parte notable del sentido físico de la transformación de bloques de espín debe encontrarse en el modo cómo cambian los acoplamientos entre espines. Las reglas para deducir los nuevos acoplamientos a partir de los antiguos en cada paso suelen ser complicadas, pero el efecto del cambio puede ilustrarse con un ejemplo bastante simple. Aunque las hipótesis no son realistas, discutiré un modelo en el que no se introducen acoplamientos con un alcance mayor que el original entre vecinos inmediatos. El único cambio en el acoplamiento es un ajuste en el valor de *K*, que equivale a un desplazamiento en la temperatura. Este cambio de *K* tendrá además una forma sencilla: en cada paso del método, la fuerza de acoplamiento en la nueva red valdrá el cuadrado del acoplamiento de la red antigua. Si llamamos K' al nuevo acoplamiento, vendrá dado por la ecuación $K' = K^2$.

Supongamos que, en un estado inicial, *K* es igual a 1/2 (que quiere decir que la temperatura tiene un valor inicial de 2 en las unidades arbitrarias que se emplean aquí). En la red menos tupida, obtenida como resultado de la transformación de bloques de espín, *K* se sustituirá por K', con un valor de $(1/2)^2$, es decir 1/4. La repetición de la transformación proporciona valores sucesivos de 1/16, 1/256, etcétera, en una serie que tiende rápidamente a cero. Con cada iteración, el sistema de espines se convierte en un nuevo sistema que, además de poseer una red más ancha, ofrece un acoplamiento más débil entre los espines. Como *K* vale $1/T$, la temperatura aumenta con cada iteración y la red tiende al límite de temperatura infinita y espines al azar.

Si la fuerza de acoplamiento inicial se hace igual a 2 (de manera que la temperatura tiene el valor 1/2), el acoplamiento aumenta en cada paso del cálculo. Después de la primera transformación de bloques de espín, la fuerza de acoplamiento es 4, luego 16, 256; por último, el acoplamiento se hace infinito. Al mismo tiempo, desde luego, la temperatura baja y el sistema se aproxima al estado de temperatura cero, en el que todos los espines están alineados.

Debe recalcarse que lo que se analiza no es la evolución de ningún sistema de espines cuando cambia la temperatura. No se calienta ni enfría nada. En cambio, en cada paso se crea un nuevo sistema de espines, un sistema caracterizado por un conjunto distinto de acoplamientos entre los espines. El comportamiento a gran escala o largo alcance de la nueva red equivale al comportamiento que se observaría en la red original a una temperatura distinta.

Hay un valor inicial de *K* para el cual esta sucesión no tiende ni a infinito ni a cero; nos referimos al valor $K = 1$. Como 1^2 es igual a 1, K' permanece igual a *K*, independientemente de cuantas veces se repita la transformación. Cuando *K* es igual a 1, el sistema se dice que está en un punto fijo, donde la aplicación continuada de la transformación del grupo de renormalización deja inalteradas todas las propiedades esenciales de la red. De hecho, los valores $K = 0$ y $K =$ infinito representan puntos fijos, ya que cero al cuadrado es también cero e infinito al cuadrado es también infinito. Cero e infinito se consideran, sin embargo, puntos fijos triviales, mientras que el valor $K = 1$ corresponde al punto crítico.



PENDIENTE de la superficie de los parámetros en la vecindad del punto fijo crítico. La pendiente determina las propiedades macroscópicas del modelo de Ising. Si se representan las trayectorias para muchos valores iniciales de *K* cerca del valor crítico (que en este caso es $K = 1$), la pendiente en el punto de silla determina la velocidad con la que las trayectorias se alejan hacia los puntos fijos triviales $K = 0$ y $K =$ infinito. Si la superficie es, en comparación, plana (*izquierda*), una trayectoria con un valor inicial de *K* tal como $K = 1{,}01$ pasa cerca del punto de silla. Cuando la superficie está más ondulada (*derecha*), la correspondiente trayectoria se curva más bruscamente hacia la sima. Por ser la temperatura el inverso de *K*, la pendiente cerca del punto fijo revela cómo cambian las propiedades del sistema cuando los valores de la temperatura se alejan de la temperatura crítica.

En esta discusión de la técnica de los bloques de espines todos los efectos de la transformación se han expresado en función de un parámetro único: la fuerza de acoplamiento entre vecinos inmediatos, *K*. De hecho, la transformación introduce otros muchos parámetros, correspondientes cada uno a un acoplamiento de más largo alcance. Todas las posibles combinaciones de estos parámetros pueden representarse geométricamente construyendo un espacio multidimensional imaginario en el que la distancia medida a lo largo de cada dimensión corresponde a una variación en uno de los parámetros. Cada estado inicial del sistema de espines y cada transformación de bloques de espín del mismo puede representarse por un punto en una superficie contenida en este espacio de los parámetros.

En la descripción geométrica del método del grupo de renormalización, se pone de manifiesto la importancia de los puntos fijos. Para el sistema bidimensional de Ising, la superficie en el espacio de los parámetros tiene la forma de un paisaje ondulado con dos picos agudos y dos profundas simas. La cresta que conecta los picos y la depresión que conecta las simas se cruzan en el centro en un punto en silla de montar [*véase la ilustración de la página anterior*]. Una sima es el punto fijo $K = 0$; la otra es el punto fijo $K =$ infinito. El punto fijo crítico cae en el punto de equilibrio precario de la silla.

La transformación del sistema de un estado al siguiente puede representarse por el movimiento de una canica que corre sobre la superficie. Uno puede imaginarse una toma periódica de imágenes que recojan la posición de la canica en intervalos de un segundo; cada toma mostraría el efecto de una iteración de la transformación de bloques de espín. La transformación permite que la canica se mueva, pero la velocidad y dirección de la canica se determinan completamente por la pendiente de la superficie en cada punto que cruza.

Supongamos que la canica está colocada inicialmente cerca de la cumbre de una colina y justo a un lado de la cresta. Al principio se moverá rápidamente, porque la colina tiene mucha pendiente cerca de la cumbre, y avanzará en la dirección general del punto de silla. Cuando la canica se acerca a la silla, la pendiente se hace más suave y la canica se frena, pero nunca se para por completo. Además, como empezó a un lado de la cresta no alcanza de pleno el punto de silla; por el contrario, se desvía a un lado y empieza a acelerarse de nuevo esta vez hacia una sima.

La trayectoria de la canica describe el camino seguido por el punto que representa el sistema de espines de Ising cuando éste se transforma repetidamente por el método de bloques de espín. La posición inicial justo al lado de la cresta corresponde a un valor inicial de los varios parámetros de acoplamiento que equivale a una temperatura justo encima o justo debajo de la temperatura crítica. En términos del ejemplo simplificado descrito antes, con un solo parámetro, el valor de *K* es o bien ligeramente mayor o menor que 1. Tomar la fuerza de acoplamiento igual a 1 equivale a colocar la canica exactamente en la cresta. Se moverá luego directamente hacia el punto de silla, o punto fijo crítico. De nuevo el movimiento es rápido al principio pero se hace más lento cuando se acerca a la silla. En este caso, sin embargo, la canica permanece basculante entre las dos pendientes descendentes. Incluso después de un gran número de iteraciones permanece en el punto fijo.

Una trayectoria en la superficie en forma de silla, en el espacio de los parámetros, puede hacerse tan próxima al punto fijo como se quiera tomando el valor inicial de *K* suficientemente cerca del punto crítico. En el ejemplo considerado aquí, en el que el valor crítico de *K* es 1, el valor inicial de *K* puede ser 0,9999, que puede elevarse al cuadrado varias veces antes de que cambie apreciablemente. De ello resulta que la trayectoria pasa muy cerca del punto fijo crítico antes de que se desvíe hacia la sima de alta temperatura.

Mediante el análisis de muchas de

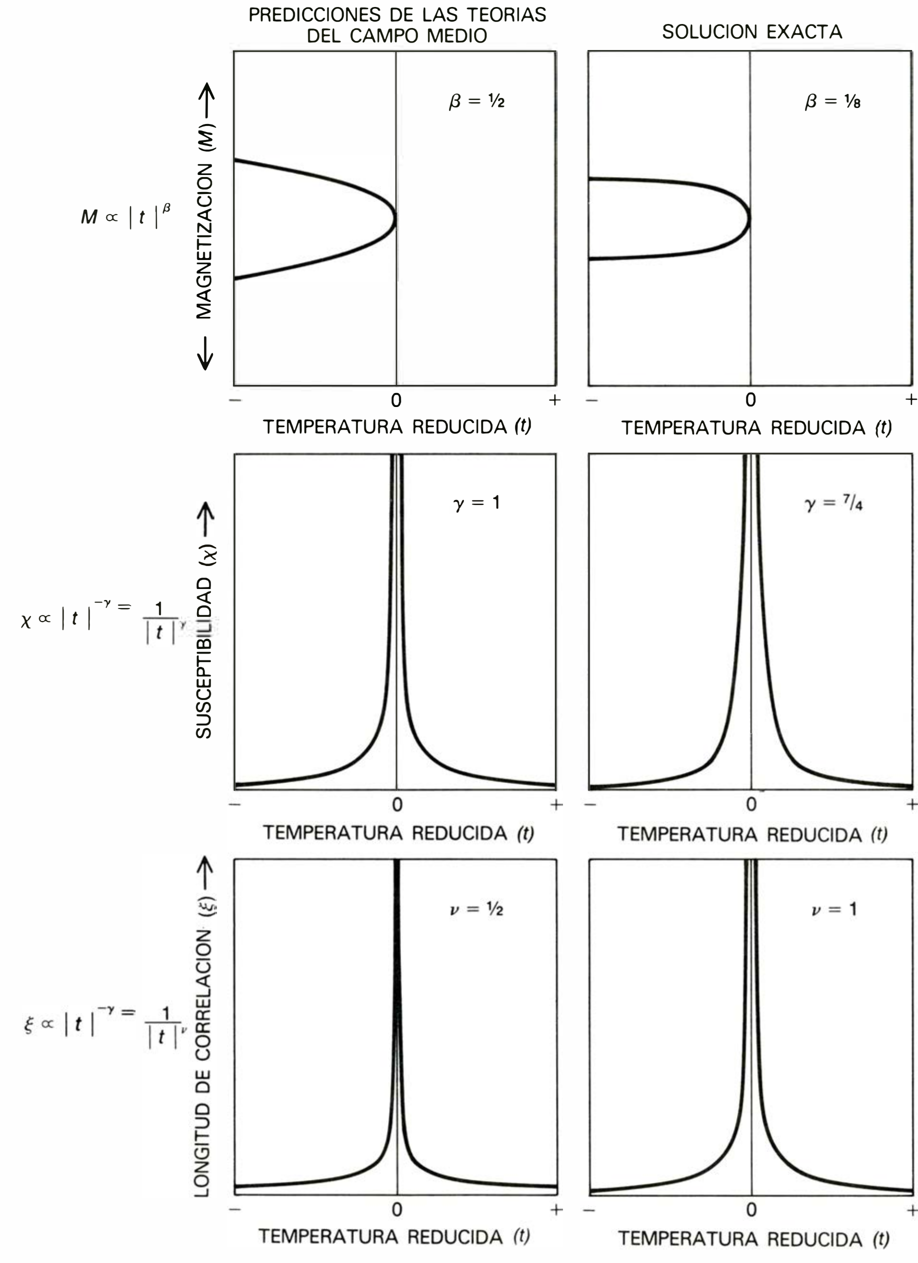
estas trayectorias puede representarse la topografía de la superficie en una pequeña región alrededor del punto de silla. La pendiente de la superficie es lo que determina cómo el sistema se aproxima al punto fijo y cómo se aleja de él. Entonces, conociendo la pendiente puede calcular cómo varían las propiedades del sistema cuando se cambian la temperatura y el acoplamiento iniciales. Esta es la información buscada para comprender los fenómenos críticos.

Las propiedades macroscópicas de un sistema termodinámico cerca del punto de equilibrio están determinadas por la temperatura. Más precisamente, propiedades tales como la magnetización espontánea, la susceptibilidad y la longitud de correlación son funciones de la diferencia entre la temperatura del sistema y la temperatura crítica, T_c. Por esta razón, es conveniente definir la temperatura de suerte que todos los puntos críticos sean equivalentes. Una magnitud adecuada es la temperatura reducida, t, definida como la diferencia entre la temperatura del sistema y la temperatura crítica, dividido por la temperatura crítica; t equivale por tanto a $(T - T_c) / T_c$. En una escala corriente de temperaturas, la escala de Kelvin por ejemplo, las temperaturas críticas de sistemas distintos toman diferentes, valores pero todos los puntos críticos poseen la misma temperatura reducida, igual a cero.

Todas las propiedades críticas son proporcionales al valor absoluto de la temperatura reducida elevada a alguna potencia. El problema de describir los fenómenos críticos consiste en determinar cuál es esta potencia o, en otras palabras, determinar los valores de los exponentes críticos. Por ejemplo, la magnetización M de un sistema de espines viene dada por la proporcionalidad M 8 $|t|^{\beta}$ donde β (la letra griega beta) es un exponente crítico y donde las líneas verticales indican el valor absoluto de t. La susceptibilidad magnética es proporcional a $1/|t|^{\gamma}$, donde γ (la letra griega gamma) es otro exponente. La longitud de correlación está asociada a un tercer exponente ν (la letra griega ni) por una relación de la misma forma: la longitud es proporcional a $1/|t|^{\nu}$.

Los primeros intentos de formular una descripción matemática de los fenómenos críticos fueron teorías de una clase que se llaman ahora teorías del campo medio. La primera de ellas fue introducida en 1873 por J. D. van der Waals como una explicación de los cambios de fase en los fluidos. Una teoría de las transiciones de fase magnéticas fue propuesta en 1907 por Pierre Weiss. En 1937 L. D. Landau, de la Academia de Ciencias de la Unión Soviética, propuso una formulación más general de la teoría del campo medio proporcionando así una base en la que podían discutirse muchos sistemas físicos. En todas estas teorías el estado de una partícula cualquiera viene determinado por las propiedades medias (propiedades tales como la magnetización del sistema) del material como un todo. De hecho, todas las partículas del sistema contribuyen por igual a la fuerza en cada punto, lo que equivale a suponer que las fuerzas poseen alcance infinito.

Las teorías del campo medio son satisfactorias cualitativamente. Dan cuenta de importantes características de los diagramas de fase de los fluidos y los ferroimanes, la más importante de las cuales es la existencia de un punto crítico. Sin embargo, las predicciones cuantitativas son menos satisfactorias:



LOS EXPONENTES CRITICOS expresan la dependencia de las propiedades macroscópicas según la diferencia entre la temperatura del sistema y la temperatura crítica. Lo mejor es expresar la temperatura en forma de temperatura reducida, t, que se define por la ecuación $t = (T - T_c)/T_c$. Todas las propiedades macroscópicas son entonces proporcionales al valor absoluto de t elevado a alguna potencia; la potencia es el exponente crítico para dicha propiedad. Los exponentes y las leyes en forma de potencias en los gráficos de la izquierda son los predichos por las teorías del campo medio, que ignoran todas las fluctuaciones. Los exponentes en los gráficos de la derecha se obtienen de una solución exacta del modelo bidimensional de Ising, presentada en 1944 por Lars Onsager, de la Universidad de Yale. Los exponentes muestran cómo las propiedades del sistema cambian cuando la temperatura o la fuerza de acoplamiento varía; ésta es la misma información que contiene la pendiente de la superficie en el espacio de los parámetros cerca del punto fijo crítico. Los exponentes pueden determinarse a partir de la pendiente, y los cálculos para el modelo bidimensional dan valores cercanos a los de Onsager.

las teorías dan valores incorrectos de los exponentes críticos. Para β, el exponente que gobierna la magnetización espontánea, la teoría del campo medio predice un valor de $1/2$; en otras palabras, la magnetización varía como la raíz cuadrada de la temperatura reducida. El exponente asociado con la suceptibilidad, γ, toma un valor 1, de manera que la susceptibilidad es proporcional a $1/|t|$. El exponente para la longitud de correlación, ν, es $1/2$, de manera que esta cantidad es proporcional a $1/\sqrt{|t|}$.

Los exponentes calculados a partir de la teoría del campo medio sugieren una forma plausible para cada una de estas funciones. La magnetización tiene dos posibles valores ($+\sqrt{t}$ y $-\sqrt{t}$) a toda temperatura por debajo del punto crítico y se anula por encima de la temperatura crítica. Tanto la susceptibilidad como la longitud de correlación tienden hacia infinito cuando t se acerca a cero tanto por el lado positivo como por el negativo. Sin embargo, se sabe que los valores de los exponentes del campo medio son incorrectos.

Para el modelo bidimensional de Ising los exponentes críticos se conocen exactamente a partir de la solución de Onsager. Los valores correctos son $\beta = 1/8$, $\gamma = 7/4$ y $\nu = 1$, que difieren significativamente de las predicciones de la teoría del campo medio e implican que el sistema tiene un comportamiento bastante distinto. Por ejemplo, la magnetización no es proporcional a la raíz cuadrada de la temperatura reducida t, sino a la raíz octava de t. De manera parecida, la susceptibilidad no se da por el recíproco de t sino por el de t elevado a la potencia 1,75, que hace que la divergencia cerca del punto crítico sea más pronunciada y abrupta.

No es difícil dar con la razón del fallo cuantitativo de las teorías del campo medio. El alcance infinito que se asigna a las fuerzas no significa siquiera una buena aproximación a la verdad. No todos los espines contribuyen por igual; los vecinos más inmediatos son mucho más importantes que los otros. La misma objeción puede expresarse de otra manera: las teorías no son capaces de contener fluctuaciones en las orientaciones de los espines o en la densidad del fluido.

En un cálculo del grupo de renormalización, los exponentes críticos se determinan a partir de la pendiente de la superficie de los parámetros en la vecindad del punto fijo. Una pendiente es simplemente una representación gráfica de un ritmo de cambio; la pendiente cerca del punto fijo determina el ritmo con que las propiedades del sistema cambian cuando la temperatura (o la fuerza de acoplamiento) varía en un intervalo estrecho cerca de la temperatura crítica. La misión de los exponentes críticos consiste también en describir el cambio del sistema en función de la temperatura; es razonable, pues, que exista una conexión entre los exponentes y la pendiente.

Los cálculos del grupo de renormalización para el sistema bidimensional de Ising han sido realizados por varios autores. En 1973 Niemeijer y van Leeuwen emplearon un método de bloques de espines para estudiar las propiedades de un sistema de espines de Ising construido en una red triangular. Yo mismo he aplicado una técnica algo distinta del grupo de renormalización, llamada de diezmado de espines ("spin decimation"), a una red cuadrada. En el diezmado de espines, en vez de agrupar en bloques de unos pocos espines cada uno, algunos espines de la red se mantienen fijos mientras se calcula una distribución de probabilidad para los espines restantes. Estos cálculos eran mucho más elaborados que el cálculo del modelo descrito antes; en mi propio trabajo, por ejemplo, se incluyeron 217 acoplamientos entre espines. Los exponentes críticos obtenidos en este cálculo están de acuerdo con los valores de Onsager en un 0,2 por ciento aproximadamente.

Debido a que se conoce una solución exacta para el modelo bidimensional de Ising, la aplicación al mismo del grupo de renormalización no es más que un ejercicio académico. Para un sistema de espines de Ising en una red tridimensional, sin embargo, no se conoce una solución exacta. Cyril Domb, del University College de Londres, y muchos otros han ideado un método para hallar valores aproximados de los exponentes

CL. UNIVERSALIDAD		MODELO TEORICO	SISTEMA FISICO	PARAMETRO ORDEN
$d = 2$	$n = 1$	Modelo de Ising bidimensional	Películas adsorbidas	Densidad superficial
	$n = 2$	Modelo XY en dos dimensiones	Películas de helio-4	Amplitud de la fase superfluida
	$n = 3$	Modelo de Heisenberg bidimensional		Magnetización
$d > 2$	$n = \infty$	Modelo "esférico"	Ninguno	
$d = 3$	$n = 0$	Camino aleatorio sin intersecciones	Conformación de polímeros de cadena larga	Densidad de finales de cadena
	$n = 1$	Modelo de Ising tridimensional	Ferroimán uniaxial	Magnetización
			Fluido cerca de un punto crítico	Diferencia de densidad entre las fases
			Mezcla cerca del punto de disolución	Diferencia de concentración
			Aleación cerca transición orden-desorden	Diferencia de concentración
	$n = 2$	Modelo de XY en tres dimensiones	Ferroimán plano	Magnetización
			Helio 4 cerca de la transición superfluida	Amplitud de la fase superfluida
	$n = 3$	Modelo de Heisenberg en tres dimensiones	Ferroimán isótropo	Magnetización
$d \leqslant 4$	$n = -2$		Ninguno	
	$n = 32$	Cromodinámica cuántica	Quarks ligados en protones, neutrones, etc.	

LA HIPOTESIS DE UNIVERSALIDAD establece que diversos sistemas físicos se comportan de la misma manera cerca de sus puntos críticos. En la mayoría de casos, los únicos factores que determinan las propiedades críticas son la dimensionalidad del espacio, d, y la dimensionalidad del parámetro de orden, n. Para sistemas magnéticos, el parámetro de orden es la magnetización, y su dimensionalidad es el número de componentes necesarias para describir el vector de espín. La mayoría de sistemas con los mismos valores de d y n son miembros de la misma clase de universalidad y poseen los mismos exponentes críticos. Por ejemplo, los ferroimanes que son parecidos al modelo tridimensional de Ising, los fluidos, las mezclas de líquidos y ciertas aleaciones son todos miembros de la clase con $d = 3$ y $n = 1$; los gráficos de sus propiedades cerca del punto crítico deben tener la misma forma. La interpretación de algunos valores de d y n es menos obvia, y valores como $n = -2$ pueden definirse matemáticamente pero no corresponden a ningún sistema físico. El modelo XY y el de Heisenberg recuerdan el modelo de Ising, pero describen ferroimanes cuyos vectores de espín tienen dos y tres componentes.

en el caso tridimensional. Primero se determinan las propiedades del sistema a alta temperatura con gran precisión y luego estas propiedades se extrapolan a la temperatura crítica. Los mejores resultados obtenidos hasta la fecha por este método dan valores de los exponentes de $\beta = 0{,}33$, $\gamma = 1{,}25$ y $\nu = 0{,}63$.

Aunque la extrapolación de la solución de alta temperatura conduce a buenas aproximaciones de los exponentes críticos, proporciona poca comprensión intuitiva de cómo se comporta el sistema cerca del punto crítico. Un cálculo del grupo de renormalización da esencialmente el mismo valor de los exponentes, y además explica importantes características universales del comportamiento crítico.

No deben olvidarse dos hechos importantes sobre los exponentes del modelo tridimensional de Ising. Se trata primero de que los valores son diferentes de los del modelo bidimensional. En las teorías del campo medio, la dimensionalidad del espacio no entra en los cálculos, y así los exponentes críticos tienen los mismos valores en cualquier espacio. La segunda sorpresa es que los exponentes no son enteros ni razones de enteros pequeños, como son en las teorías del campo medio. Pueden ser númerosos irracionales.

Si sorprende que la dimensionalidad espacial influya en los exponentes críticos, debemos destacar igualmente que ciertas otras propiedades del modelo no influyan en absoluto. Un ejemplo de un tal parámetro irrelevante es la estructura de la red. En el modelo bidimensional de Ising no importa que la red sea rectilínea, como en mi propio trabajo, o triangular, como en el modelo empleado por Niemeijer y van Leeuwen; los exponentes críticos son los mismos. Por extensión, en un ferroimán real la gran variedad de estructuras cristalinas dan todas un comportamiento crítico idéntico.

Existe una justificación intuitiva de la irrelevancia de la estructura de la red y de otras propiedades microscópicas. Un cambio en la forma de la red tiene un efecto grande sobre lo que sucede a escala del espaciado de la misma, pero su efecto disminuye cuando aumenta la escala de interés. En un cálculo del grupo de renormalización las fluctuaciones a la escala del espaciado de la red quedan promediadas después de las primeras iteraciones, de manera que modelos con muchas redes distintas tienen el mismo comportamiento crítico. A través del grupo de renormalización se ve cómo la aparición de los mismos exponentes críticos en muchos sistemas surge de la topografía de la superficie en el espacio de los parámetros. Cada estructura reticular corresponde a una posición diferente en el espacio de los parámetros; pero a la temperatura crítica, cada red se representa por un punto en algún sitio a lo largo de la cresta. Tras repetidas transformaciones del grupo de renormalización, todos estos sistemas convergen en el mismo punto fijo, el punto de silla.

La idea de que ciertas variables resultan irrelevantes para los fenómenos críticos puede extenderse a sistemas distintos de los ferroimanes. Un fluido cerca de su punto crítico, por ejemplo, tiene las mismas propiedades que el modelo de Ising tridimensional de un ferroimán. Para comprender esta identidad hay que establecer alguna correspondencia entre las propiedades macroscópicas del fluido y las del imán. La magnetización, que es el número de espines hacia arriba menos el número de espines hacia abajo, puede identificarse con la diferencia de densidad en el fluido: la densidad de la fase líquida menos la densidad de la fase vapor. Igual que la magnetización se anula en el punto de Curie, la diferencia de densidad se hace cero en el punto crítico del fluido. Estas cantidades –la magnetización y la diferencia de densidad– constituyen los llamados parámetros de orden de sus respectivos sistemas. La susceptibilidad del imán, que es el cambio de la magnetización ante un pequeño cambio dado del campo magnético aplicado, es análoga a la compresibilidad del fluido: el cambio en la densidad que resulta de un pequeño cambio dado en la presión. Como la susceptibilidad, la compresibilidad se hace infinita en el punto crítico. El comportamiento crítico del fluido y el del modelo tridimensional de Ising son idénticos en tanto que tienen la misma superficie en el espacio de los parámetros. Los dos sistemas ostentan diferentes posiciones iniciales en la superficie, pero convergen en el mismo punto de silla y, por tanto, tienen los mismos exponentes críticos.

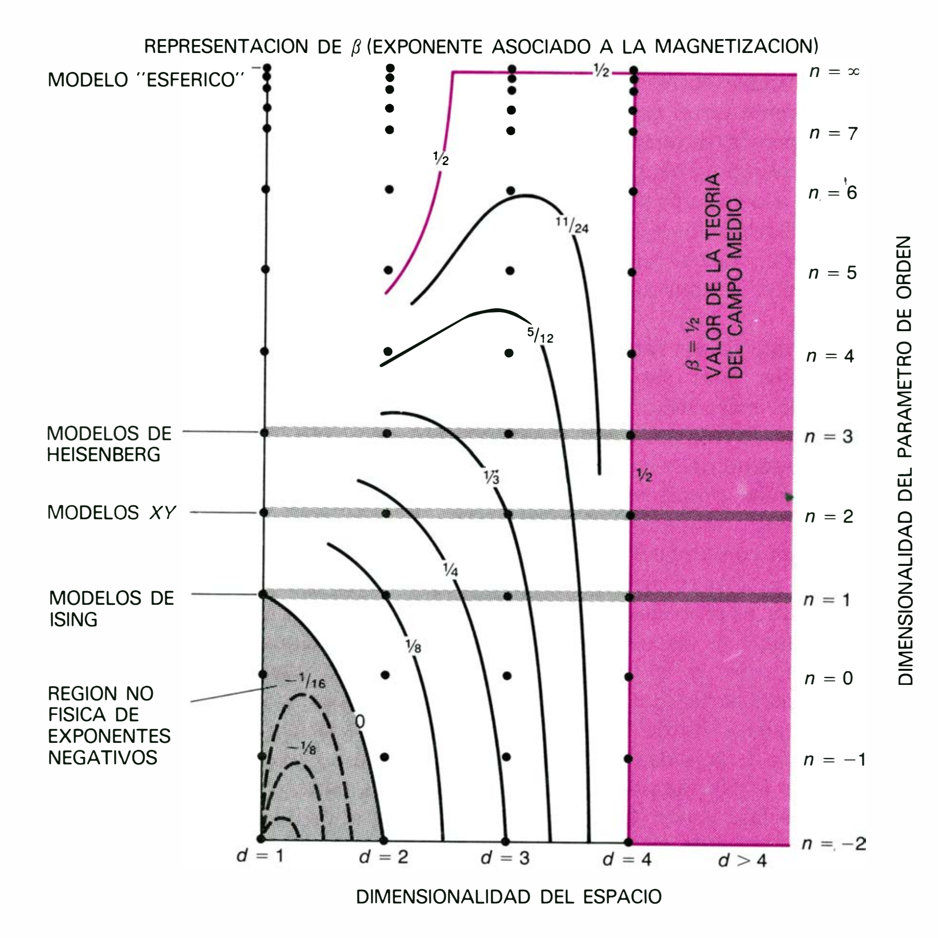
La similitud observada en el comportamiento crítico de fluidos y de ferroimanes es un ejemplo de una hipótesis más general llamada universalidad del punto crítico. Según esta hipótesis sólo dos cantidades determinan el comportamiento crítico de la mayoría de sistemas: la dimensionalidad del espacio y la dimensionalidad del parámetro de orden. Estas cantidades se denominan respectivamente d y n. Se cree que todos los sistemas que tienen los mismos valores de d y n poseen la misma superficie en el espacio de los parámetros y los mismos exponentes críticos. Y se dice que son miembros de la misma clase de universalidad.

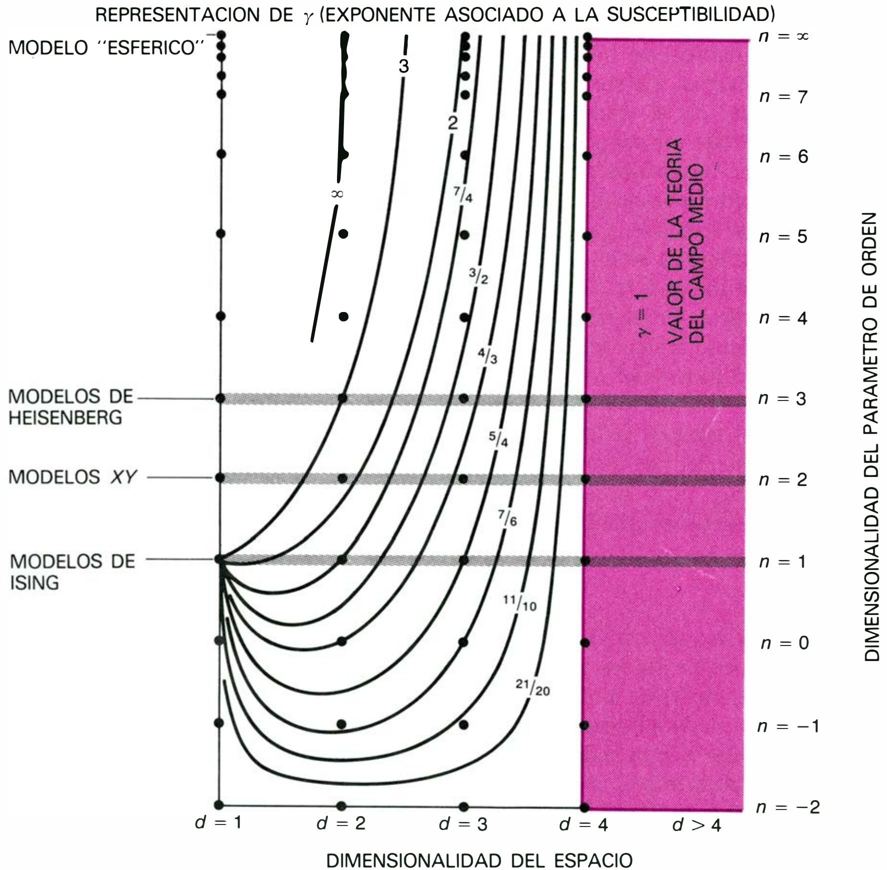
Suele ser fácil determinar la dimensionalidad del espacio, pero la dimensionalidad del parámetro de orden requiere una consideración más cuidadosa. En sistemas magnéticos, en los que el parámetro de orden es la magnetización, n es el número de componentes necesarias para definir el vector de espín. El vector de un espín de Ising puede orientarse sólo a lo largo de un eje, y así tiene sólo una componente; para el modelo de Ising n es igual a 1. Un vector de espín que puede apuntar en cualquier dirección de un plano tiene dos componentes, que se suelen representar a lo largo de los dos ejes que definen un plano. Similarmente, un vector que puede apuntar a cualquier parte del espacio tridimensional tiene tres componentes, y así n es igual a 3.

Para el modelo tridimensional de Ising d es igual a 3 y n es igual a 1. Los fluidos ordinarios pertenecen a la misma clase de universalidad. El espacio en el que los fluidos existen claramente tiene tres dimensiones. El parámetro de orden –la diferencia de densidad entre las fases líquida y de vapor– es una cantidad que tiene sólo su valor y por tanto sólo una componente; puede expresarse como un simple número, igual que el valor que puede tomar un espín de Ising.

Varios otros sistemas físicos son miembros de esta clase. Una mezcla de dos líquidos tales como aceite y agua ofrecen un comportamiento crítico cerca de la temperatura en que los fluidos componentes se hacen completamente miscibles entre sí, una temperatura llamada punto crítico de disolución. A temperaturas por debajo del punto crítico de disolución la mezcla se separa en dos fases, y el parámetro de orden se define como la diferencia de concentración entre las dos fases, otra cantidad que puede expresarse como un simple número. Algunas aleaciones, el latón por ejemplo, tienen una transición entre una fase ordenada, en la que los dos metales ocupan puestos alternados en una red regular, y una fase desordenada, en la que su distribución es menos uniforme. El parámetro de orden en este sistema es de nuevo una diferencia de concentración, de manera que n es igual a 1. Se espera que todos estos sistemas tengan los mismos exponentes críticos que el modelo tridimensional de Ising. Así sucede con algunos ferroimanes reales, aquellos que sólo se imantan fácilmente según un eje. La evidencia experimental disponible confirma estas predicciones.

La hipótesis de universalidad sería





VARIACION DE LOS EXPONENTES CRITICOS con la dimensionalidad del espacio (*d*) y del parámetro de orden (*n*). Ello sugiere que los sistemas físicos de clases de universalidad distintas deben tener propiedades críticas distintas. Los exponentes pueden calcularse como funciones continuas de *d* y *n*, pero sólo son físicamente posibles sistemas con un número entero de dimensiones. En un espacio de cuatro o más dimensiones, todos los exponentes críticos toman los valores predichos por las teorías del campo medio. Los gráficos fueron preparados por Michael E. Fisher, de la Universidad de Cornell.

trivial si los exponentes críticos fueran números enteros o fracciones simples tales como 1/2. Muchas leyes físicas tienen estos exponentes y no hay una razón fuerte para postular una conexión entre ellas. Tanto la gravitación como el electromagnetismo tienen una ley que depende del inverso al cuadrado (un exponente de −2), pero esta coincidencia no demuestra que las dos fuerzas sean iguales. La correspondencia de exponentes parece remarcable, sin embargo, cuando los valores no son números redondos sino fracciones tales como 0,63. La convergencia de muchos sistemas con estos valores no puede ser una coincidencia. Es evidente que todos los detalles de estructura física que distinguen un fluido de un imán son menos importantes que las propiedades geométricas expresadas por los valores de d y n.

El modelo bidimensional de Ising ($d = 2$, $n = 1$) tipifica una clase de sistemas que están confinados a un espacio de dos dimensiones. Un ejemplo es una película delgada de líquido; otro es un gas adsorbido en una superficie sólida. Un ferroimán ordinario pertenece a la clase de $d = 3$ y $n = 3$, es decir, la red es tridimensional y cada espín tiene tres componentes, de manera que puede apuntar en cualquier dirección. Cuando los espines se restringen a estar en un plano, la clase se reduce a $d = 3$ y $n = 2$. En esta misma clase están la transición superfluida del helio 4 líquido y las transiciones superconductoras de varios metales.

Otras clases de universalidad tienen valores de d y n cuya interpretación es algo menos obvia. El caso de $d = 4$ es de interés en física de partículas elementales, donde una de las cuatro dimensiones corresponde al eje del tiempo. En una red de espines teórica llamada el modelo esférico, en la que un espín individual puede tener cualquier magnitud y sólo se restringe el valor de la suma de todos los spines, n es efectivamente infinita. Un camino al azar, que no pase dos veces por un mismo punto, a través de los puntos de la red, describe la representación en el espacio de un polímero de cadena larga; Pierre Gilles de Gennes, del Collège de France, ha demostrado que este problema pertenece a una clase de universalidad con n igual a cero. En modelos teóricos n puede incluso tomar el valor −2, aunque el sentido físico de un número negativo de componentes de un vector no es claro.

Los únicos valores de d y n que tienen un sentido físico directo son los valores enteros. Esto resulta especialmente claro en el caso de d, puesto que un espacio

con un número no entero de dimensiones es incluso difícil de imaginar. En los cálculos del grupo de renormalización, sin embargo, *d* y *n* aparecen en ecuaciones donde se les permite variar continuamente en algún intervalo. Incluso es posible dibujar un gráfico en el que los valores de los exponentes críticos se representen como funciones continuas de *d* y *n*. Los exponentes tienen valores bien definidos no sólo para dimensiones enteras sino también para todas las dimensiones fraccionarias entre los enteros. Tal gráfico muestra que los exponentes se aproximan a los valores dados por la teoría del campo medio cuando el número de dimensiones espaciales se acerca a 4. Cuando *d* es exactamente igual a 4, y para todos los valores mayores de *d*, los resultados del campo medio son exactos. Esta observación ha dado lugar a un método importante de realizar cálculos del grupo de renormalización. La dimensionalidad del espacio se expresa como igual a $4-\varepsilon$, donde ε (la letra griega épsilon) representa un número que se supone pequeño. Los exponentes críticos pueden calcularse entonces como la suma de una serie infinita de términos que incluyen potencias progresivamente mayores de ε. Si ε es menor que 1, una potencia grande de ε tendrá un valor pequeño, y puede obtenerse una precisión razonable despreciando todos los términos de la serie infinita excepto los primeros.

Este método de cálculo, que se llama el desarrollo en épsilon, ha sido investigado por Michael E. Fisher, de la Universidad de Cornell, y por el autor. Se trata de un método general para resolver todos los problemas a los que puede aplicarse la teoría del campo medio y representa el sucesor natural de la teoría de Landau. Realmente proporciona respuestas en forma de correlaciones a los valores dados por la teoría del campo medio. El método de bloques de espines es la técnica más transparente, pero el desarrollo en épsilon es la más potente.

No es del todo sorprendente que los exponentes críticos deban converger a los valores del campo medio cuando el número de dimensiones espaciales aumenta. La hipótesis fundamental de las teorías del campo medio es que la fuerza en cada punto de la red está influenciada por las condiciones en muchos otros puntos. El número de puntos vecinos inmediatos aumenta junto con el número de dimensiones espaciales. En una red unidimensional, cada punto tiene sólo dos vecinos inmediatos; en una red bidimensional, cuatro; en una red tridimensional, seis y, en una red cuadridimensional, ocho. Por tanto, a medida que la dimensionalidad aumenta, la situación física empieza a parecerse más a la hipótesis básica de la teoría del campo medio. Sigue siendo un misterio, sin embargo, por qué $d=4$ debe marcar una frontera brusca por encima de la cual los exponentes del campo medio son exactos. He discutido principalmente las aplicaciones del grupo de renormalización a los fenómenos críticos. Sin embargo, la técnica no se limita a estos problemas y de hecho no empezó con ellos.

El método llamado renormalización fue inventado en los años cuarenta como parte del desarrollo de la electrodinámica cuántica, la teoría moderna de las interacciones entre las partículas cargadas eléctricamente y el campo electromagnético. La dificultad encontrada en la formulación de la teoría puede entenderse como debida a las múltiples escalas de longitud. Por algún tiempo había parecido que la carga del electrón predicha por las teorías mecánico-cuánticas era infinita, una predicción que estaba en serio conflicto con la carga medida. La teoría renormalizada de la electrodinámica no suprime el infinito; por el contrario, el electrón se define como una partícula puntual cuya carga "desnuda" es infinita. Sin embargo, en la electrodinámica cuántica la carga desnuda tiene el efecto de inducir una carga de polaridad opuesta en el vacío circundante, que cancela la mayor parte del infinito, dejando sólo la pequeña carga neta que se observa en los experimentos ordinarios.

Imaginemos una partícula de prueba que pueda medir la carga del electrón en una zona arbitrariamente cercana. En una zona grande encontraría el valor finito familiar, que es la diferencia entre la carga desnuda y la carga inducida. Conforme fuese penetrando las capas de apantallamiento, la carga medida aumentaría y, cuando la zona se redujese a cero, la carga se haría infinita. El método de la renormalización proporciona un medio para sustraer la carga infinita apantallante de la carga desnuda infinita, para así obtener una diferencia finita.

En los años 50 varios físicos, entre los que estaba Murray Gell-Mann y Francis E. Low, indicaron que el método de la renormalización usado en la electrodinámica cuántica no era único. Propusieron una formulación más general, que es la versión original del grupo de renormalización. En su aplicación del método a la electrodinámica cuántica se construye una expresión matemática que da la magnitud de la carga a alguna distancia determinada del electrón. Se analiza luego la forma de la expresión cuando la distancia a la que se saca la medida se hace aproximar a su valor límite de cero. La arbitrariedad del método está en la elección de la distancia inicial. Puede seleccionarse cualquier valor sin que cambien los resultados finales, de manera que hay un conjunto infinito de métodos de renormalización equivalentes.

Un "grupo" en matemática es un conjunto de transformaciones que cumple un requerimiento especial: el producto de cualquier par de transformaciones debe también ser un miembro del conjunto. Por ejemplo, las rotaciones son transformaciones que constituyen un grupo, ya que el producto de cualquier par de rotaciones es también una rotación. Ello significa, en el caso del grupo de renormalización, que el procedimiento puede iterarse indefinidamente, ya que aplicar el procedimiento dos veces es equivalente a aplicar el producto de las dos transformaciones. Hablando con propiedad, el grupo de renormalización no es más que un semigrupo, porque el inverso de una transformación no está definido. La razón de esto puede verse simplemente en la técnica de bloques de espines aplicada al modelo bidimensional de Ising. Un bloque de nueve espines puede condensarse en un único espín promedio, pero la configuración de espines original no puede reconstruirse a partir del promedio, ya que se ha perdido información esencial.

La versión del grupo de renormalización indicada en este artículo difiere en varios aspectos de la introducida por Gell-Mann y Low. La primitiva versión de la técnica es sólo útil para entender problemas que pueden resolverse por uno de los métodos tradicionales de la física: encontrando alguna expresión aproximada para el comportamiento del sistema y calculando entonces aproximaciones mejores como una serie de perturbaciones a partir de la expresión original. Además, en la formulación original sólo se permite que varíe una cantidad; en el ejemplo indicado antes es la carga del electrón. En consecuencia, la superficie en el espacio de los parámetros no es un paisaje multidimensional sino una simple línea. La versión moderna del grupo de renormalización, introducida por el autor en 1971, da acceso a un espectro mucho más amplio de problemas físicos. Y lo que es igualmente importante, da un significado físico al método de la renormalización, que de otra manera parece puramente formal.

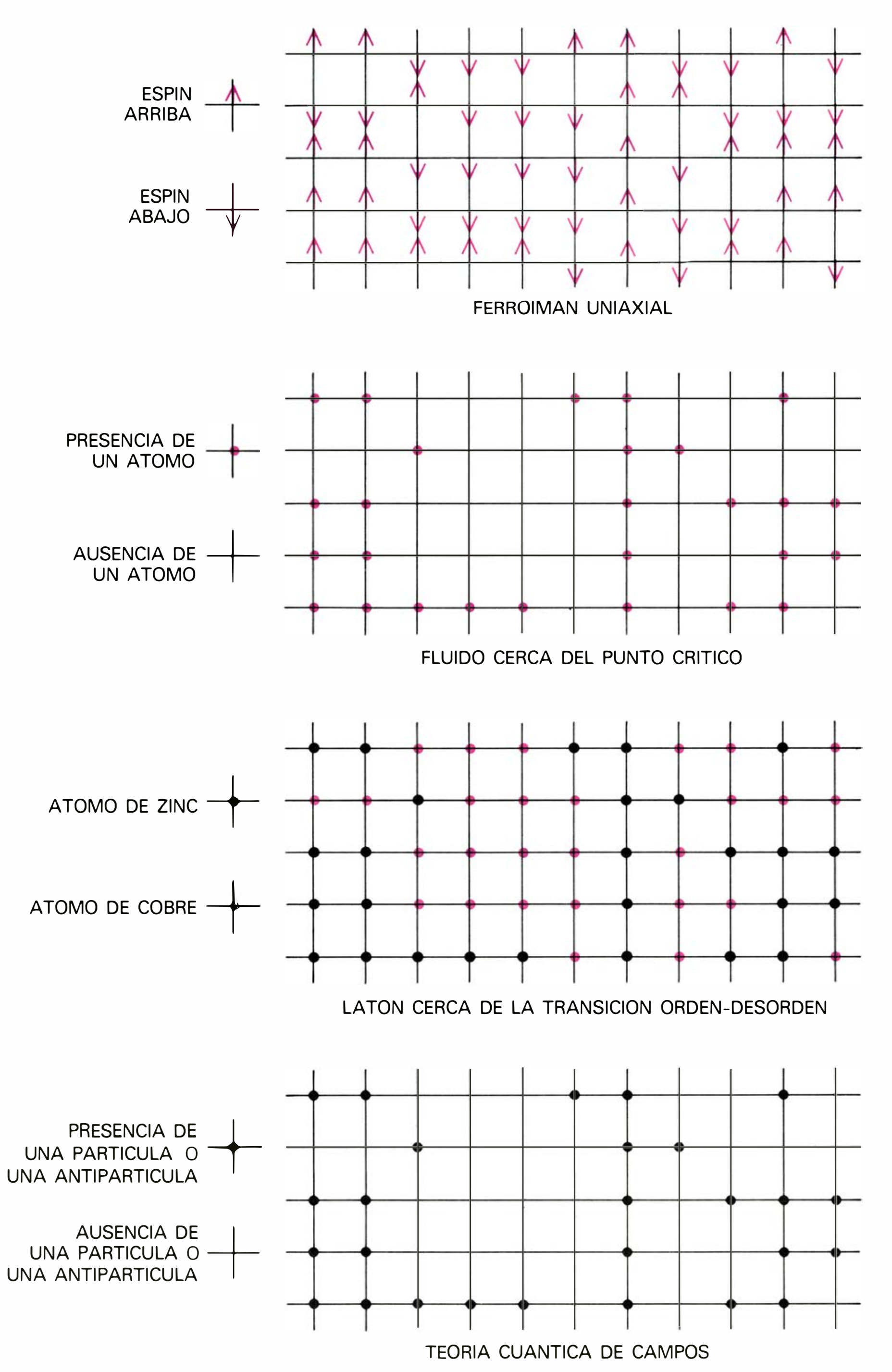
En los últimos años hemos intentado

aplicar la nueva versión del grupo de renormalización a un problema de la física de las partículas elementales. El problema estriba en cómo describir las interacciones de los quarks, las partículas elementales hipotéticas que se cree que forman protones, neutrones y una multitud de partículas relacionadas. En cierto sentido, el problema se parece estrechamente al problema de la renormalización de la electrodinámica cuántica; en otro sentido, es justo lo opuesto.

En la electrodinámica cuántica la carga del electrón aumenta cuando uno se aproxima más al electrón. Para las interacciones de los quarks, la propiedad análoga a la carga eléctrica se llama color, y por esta razón la teoría de las interacciones de los quarks se denomina cromodinámica cuántica. Cuando la carga de color de un quark se mide de cerca, parece disminuir conforme la distancia se hace más pequeña. En consecuencia, dos quarks que estén muy juntos casi no interaccionarán: el acoplamiento entre ellos será débil. Por contra, cuando los quarks se separen, la carga de color efectiva aumentará y quedarán fuertemente ligados. Mientras que un electrón induce una carga compensadora en el espacio circundante, un quark parece inducir una carga de color de la misma polaridad, que aumenta su propia carga a larga distancia. Es una hipótesis ampliamente aceptada que el acoplamiento efectivo entre los quarks aumenta sin límite cuando la distancia entre ellos sobrepasa el diámetro de un protón, que es de unos 10^{-13} centímetros. Si esto es verdad, un quark podría ser arrancado de un protón sólo a base de gastar una cantidad infinita de energía. Los quarks estarían permanentemente confinados [véase "El modelo en bolsa del confinamiento de los quarks", por Kenneth A. Johnson, INVESTIGACION Y CIENCIA, septiembre, 1979].

Una manera de visualizar el enlace de los quarks es construyendo líneas de fuerza imaginarias entre ellos. La fuerza del acoplamiento es entonces proporcional al número de líneas por unidad de área que cruzan cualquier superficie entre las partículas. En el caso de los electrones, cuando las partículas están separadas, las líneas de fuerza se expanden en el espacio, y así hay menos líneas por unidad de área. La densidad de las líneas disminuye como el cuadrado de la separación, lo que proporciona la familiar ley del cuadrado inverso de la fuerza electromagnética. Con los quarks, por otro lado, la hipótesis predominante hace que las líneas de fuerza no se expandan en el espacio; permanecen confinadas en un delgado tubo, o cuerda, que une directamente los quarks. En consecuencia, el número de líneas por unidad de área permanece constante, con independencia de cuál sea la distancia, y los quarks no pueden separarse. Aunque esta explicación del confinamiento de los quarks tiene un interés a nivel intuitivo, es sólo una explicación cualitativa. Nadie ha logrado por ahora deducir el confinamiento de los quarks a partir de la teoría básica de la cromodinámica cuántica.

El confinamiento de los quarks tiene muchas escalas de longitud y de energía y, por tanto, es un candidato a los



UN SISTEMA RETICULAR puede interpretarse como un modelo no sólo de un ferroimán sino también de otros sistemas físicos que tienen fluctuaciones de muchas escalas. El modelo de Ising describe un ferroimán uniaxial, con un eje de magnetización privilegiado. Puede aplicarse también a un fluido cerca de su punto crítico, en el que cada posición de la red está ocupada por un átomo o está vacía, de suerte que las fluctuaciones son variaciones de densidad. Una aleación como la del latón tiene una estructura similar, en la que cada lugar está ocupado por una clase u otra de metal. En todos estos sistemas, las fluctuaciones son térmicas; en las teorías cuánticas de campos que describen las interacciones de las partículas hay fluctuaciones cuánticas del vacío, que permiten que partículas y antipartículas aparezcan espontáneamente. Puede formularse una teoría cuántica de campos en una red especificando que las partículas y antipartículas pueden crearse y aniquilarse sólo en los puntos de la red.

métodos del grupo de renormalización. Yo mismo he formulado una versión del problema en el que los quarks ocupan las posiciones de una red en una configuración espacio-tiempo cuadridimensional y en el que están conectados por "cuerdas" que siguen las líneas que conectan los puntos de la red. La red es una estructura estrictamente artificial, sin análogo en el espacio-tiempo real, y debe finalmente desaparecer de la teoría. Ello puede lograrse haciendo que el espaciado de la red tienda a cero.

Como en el estudio de los sistemas ferromagnéticos, una transformación del grupo de renormalización se aplica repetidamente a la red de quarks y cuerdas. De esta manera, la interacción de los quarks puede analizarse en separaciones progresivamente mayores. La cuestión que debe contestarse es si las líneas de fuerza permanecen confinadas en fibras tubulares o si se dispersan en la red a medida que se aumenta la escala de longitud. Los cálculos están cerca del límite de las posibilidades de las computadoras digitales de la generación actual. De momento no tenemos las respuestas adecuadas.

Hay otros muchos problemas a los métodos del grupo de renormalización, si bien no han sido aún expresados de forma tal que puedan resolverse. La percolación de un fluido a través de una matriz sólida, como el agua pasando a través del suelo o el café a través de los granos de café molidos, implica agregaciones de fluido de muchas escalas. La turbulencia de los fluidos representa un problema de notoria dificultad que ha resistido más de un siglo de esfuerzos para describirlo matemáticamente. Se caracteriza por configuraciones de muchos tamaños característicos. En la atmósfera, por ejemplo, el flujo turbulento varía en escala desde los pequeños "remolinos de polvo" hasta los huracanes arrasadores.

Un problema que ha cedido ante el grupo de renormalización es un fenómeno de física del estado sólido llamado el efecto kondo, en honor del físico japonés Jun Kondo. El efecto se observa en metales no magnéticos, tales como el cobre, cuando están contaminados con una pequeña concentración de átomos magnéticos. Las teorías más simples predicen que la resistencia eléctrica de semejante metal disminuirá continuamente cuando se reduzca la temperatura. De hecho la resistencia alcanza un valor mínimo a una temperatura finita y sube luego de nuevo cuando la temperatura se reduce más. La anomalía no fue nunca de una importancia acuciante porque una explicación de la misma no daba luz a propiedades más generales de los sólidos, pero preocupó a los físicos durante más de 40 años, pareciendo estar siempre justo debajo del alcance de los métodos disponibles. La raíz de la dificultad es que los electrones de conducción en el metal pueden tener cualquier energía en un intervalo de unos pocos volt, pero las perturbaciones en esta energía son significativas por debajo de un nivel de unos 10^{-4} volt. El problema fue resuelto finalmente en 1974 cuando yo mismo terminé un cálculo del grupo de renormalización de las energías de los electrones a todas las temperaturas hasta el cero absoluto.

Debe destacarse una serie más reciente de cálculos del grupo de renormalización, notable porque se han conseguido predicciones que han confirmado directamente los experimentos. Los cálculos se refieren al modelo de red de espines en que d es igual a 2 y n es igual a 2, o, en otras palabras, se refiere a una red bidimensional de espines de dos componentes. Se ha demostrado que en este sistema de espines no es posible una fase con orden de largo alcance, pero los estudios del grupo de renormalización hechos por J. M. Kosterlitz, de la Universidad de Birmingham, y David J. Thouless, de la Universidad de Yale, han demostrado que el comportamiento del sistema cambia bruscamente a una temperatura crítica. Estos hallazgos se han aplicado a estudios de películas delgadas de helio 4 superfluido, que también pertenecen a la clase de universalidad de $d = 2$ y $n = 2$. En particular, Kosterlitz y David R. Nelson, de la Universidad de Harvard, han predicho un salto discontinuo en la densidad de la fracción superfluida de la película. Desde entonces tal salto ha sido observado experimentalmente por John D. Reppy, de Cornell, y otros; se ha encontrado que tiene la magnitud predicha.

Por todo el trabajo que se ha dedicado al grupo de renormalización podría parecer que los resultados obtenidos hasta ahora no resultan suficientes. Debe tenerse en cuenta que los problemas a los que se aplica el método son algunos de los problemas más difíciles conocidos en las ciencias físicas. Si no lo fueran, habrían sido resueltos hace tiempo por métodos más fáciles. De hecho, un número importante de los problemas no resueltos de la física deben su dificultad a una multiplicidad de escalas. La trayectoria más prometedora hacia su solución, incluso aunque sea una trayectoria ardua, es un mayor refinamiento de los métodos del grupo de renormalización.

El control hormonal del comportamiento en un lagarto

El anolis verde constituye un sujeto apropiado para examinar la manera en que las glándulas sexuales y el cerebro interactúan para orquestar el comportamiento sexual de machos y hembras

David Crews

Generaciones de niños americanos se han familiarizado con el lagarto anolis verde *(Anolis carolinensis)*, al que conocen con el nombre de camaleón, pues con esta denominación se vende en las tiendas de animales domésticos. Pero es muy reciente la atención prestada por los biólogos a este saurio, un animal óptimo para el estudio en el laboratorio de la interacción entre el comportamiento y las hormonas, que son los "mensajeros químicos" del cuerpo que actúan a cierta distancia del lugar donde se producen. Un interés temprano por el anolis verde (que de niño yo creía también que era un camaleón) me llevó más tarde, como biólogo del comportamiento, a trabajar con él al objeto de investigar las bases del comportamiento reproductor. Entre otras cosas, los descubrimientos permiten una nueva perspectiva sobre la función adaptativa de la relación entre el comportamiento y las hormonas que se advierte en animales pertenecientes a especies muy distintas.

El camaleón verdadero y el anolis verde tienen muchas cosas en común. Los dos son saurios, ambas especies viven en árboles o arbustos y se alimentan principalmente de insectos. Ambos pueden cambiar de color, aunque la capacidad del anolis para hacerlo es mucho más limitada que la del camaleón. Estos rasgos han hecho que camaleones y anolis sean animales domésticos populares. El cambio de color del anolis no está relacionado, como mucha gente cree, con el color de su entorno, sino que está determinado por factores tales como la luz y la temperatura o por emociones tales como miedo, triunfo o derrota. Los camaleones son animales del Viejo Mundo, mientras que los anolis se encuentran en las regiones más cálidas de América del Norte y del Sur. El camaleón común pone de dos a 40 huevos por puesta, el anolis un huevo único.

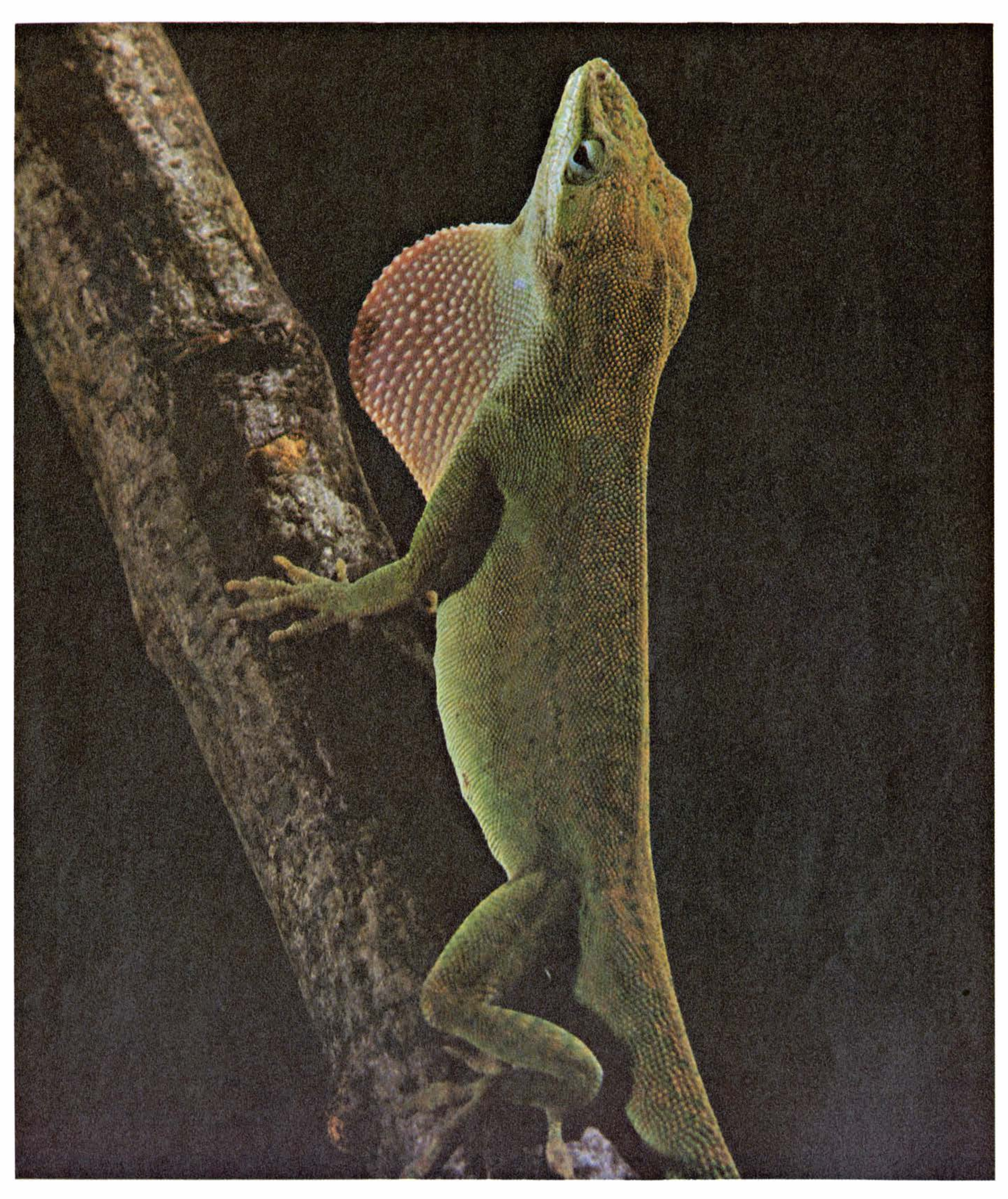
El valor particular del anolis verde como animal experimental reside en su abundancia y en que, bajo condiciones apropiadas, puede establecer en el laboratorio el mismo sistema social y el mismo comportamiento que exhibe en su ambiente natural. Una población típica puede encontrarse en Louisiana meridional. Allí los anolis son reproductivamente inactivos y permanecen esencialmente aletargados desde finales de septiembre hasta finales de enero. Durante ese período se amontonan en grupos detrás de la corteza suelta de árboles muertos o bajo piedras y troncos caídos. A finales de enero o principios de febrero, los machos aparecen y establecen territorios de reproducción. Las hembras empiezan a ser activas alrededor de un mes después; para mayo, cada hembra pone su único huevo en el suelo a intervalos de 10 a 14 días. La estación reproductora, que se extiende por todo agosto, es seguida por cerca de un mes en el que tanto machos como hembras son refractarios, es decir, insensibles a los estímulos ambientales y sociales que en primavera los indujeron a empezar la reproducción.

Los estudios sobre el comportamiento reproductor de varias especies animales, y sobre todo las investigaciones del malogrado Daniel S. Lehrman con la paloma torcaz, han demostrado que el comportamiento es la consecuencia de dos sistemas entrelazados, uno de ellos representado por el animal y su estado interno y el otro por el animal y su ambiente. Así, en muchas aves y mamíferos la duración creciente de la luz diurna en primavera actúa sobre el cerebro de los machos estimulando la glándula pituitaria para que secrete cantidades crecientes de hormonas gonadotropas. En los saurios, la clave estimuladora primaria es el aumento de la temperatura. Paul Licht, de la Universidad de California en Berkeley, ha demostrado que los lagartos difieren, además, de aves y mamíferos porque secretan una sola hormona gonadotropa y no dos.

La corriente sanguínea transporta la gonadotropina hasta los testículos, donde estimula la producción de espermatozoides y la secreción de andrógenos (término colectivo para designar a las hormonas masculinas), específicamente testosterona. La concentración creciente de testosterona en la sangre actúa, a su vez, sobre zonas específicas del cerebro para modular la secreción ulterior de gonadotropina por parte de la pituitaria y activar el comportamiento sexual del macho. El comportamiento de éste se convierte entonces en un rasgo importante del ambiente que estimula la pituitaria de la hembra para que secrete gonadotropina, que rápidamente induce el desarrollo ovárico, la secreción de hormonas esteroides y el comportamiento sexual femenino. Cuando la hembra responde al macho, su comportamiento influye sobre el de éste, e indirectamente sobre la fisiología del macho, de modo que el ciclo recomienza de nuevo.

El anolis verde despliega un rico repertorio de comportamiento. Un macho sexualmente activo suele patrullar por su territorio, deteniéndose en apostaderos prominentes para ejecutar una "exhibición de afirmación". Se caracteriza por un movimiento de asentimiento de la cabeza que es típico de la especie y está coordinado con la extensión de una papada gular roja. Si un macho de anolis verde extraño penetra en el territorio, el macho residente responde con una "exhibición de desafío". Se caracteriza por una compresión lateral extrema del

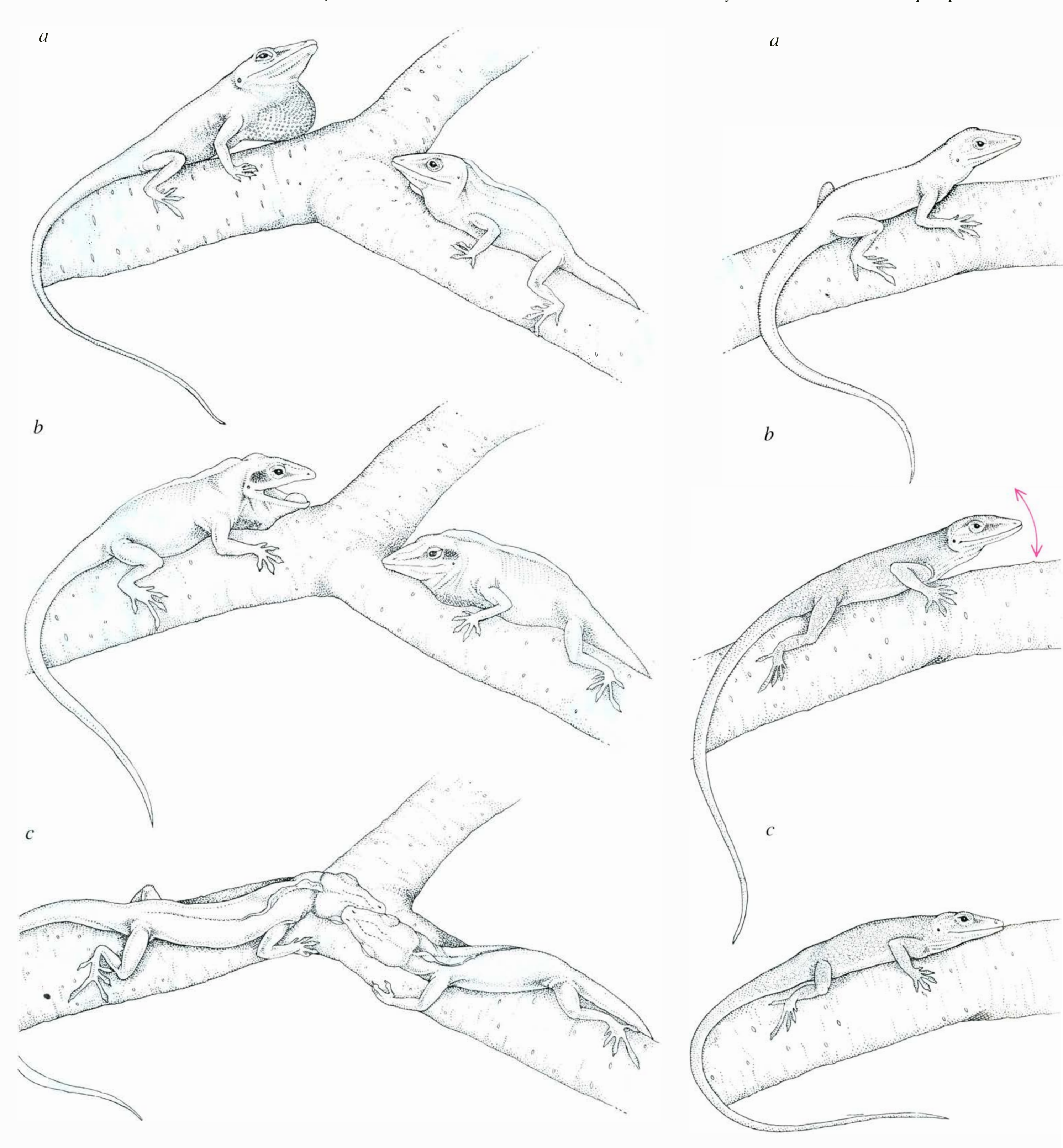
cuerpo y un movimiento de asentimiento de la cabeza muy estereotipado. Si el animal intruso no responde inmediatamente con una "exhibición de sumisión" (una sacudida rápida de la cabeza), o bien contesta con agresividad al desafío, el residente se acerca al intruso y empieza la lucha. Durante la misma, la papada no se halla extendida, pero la región gular está congestionada por haber descendido el aparato que sirve de inserción a la lengua. Al ir avanzando la lucha, a lo largo del dorso y del cuello se yergue una cresta y se forma una mancha negra inmediatamente detrás del ojo. No es raro que los machos traben sus mandíbulas mientras giran uno alrededor del otro, intentando echarse mutuamente del apostadero. Terminado el lance, el macho vencedor (casi siempre el residente) trepará con seguridad a un



EXHIBICION DE CORTEJO que realiza un lagarto anolis verde (*Anolis carolinensis*) macho que ha visto a una hembra, que no aparece en la fotografía. El macho avanza hacia la hembra, deteniéndose de vez en cuando para extender la roja papada de su garganta y para ejecutar una serie de movimientos de asentimiento. El número de reverencias varía de un macho a otro, lo que quizás ayuda a la hembra a distinguir entre varios machos.

apostadero elevado y ejecutará una serie de exhibiciones de afirmación.

En una "exhibición de cortejo" el macho avanza hacia la hembra, deteniéndose varias veces para ejecutar una serie de movimientos de asentimiento y de exhibiciones de la papada. El número de movimientos de asentimiento varía mucho de un macho a otro, y esta variación puede ayudar a las hembras a diferenciar a los distintos machos. Si la hembra es sexualmente receptiva, permitirá que el macho se le acerque y agarre por el cuello; después, el macho monta sobre la hembra y curva su cola bajo la de ella. Esta acción pone en contacto ambas regiones cloacales, de modo que el macho pueda evaginar uno de sus dos penes. (Entre los vertebrados, saurios y ofidios se caracterizan por poseer

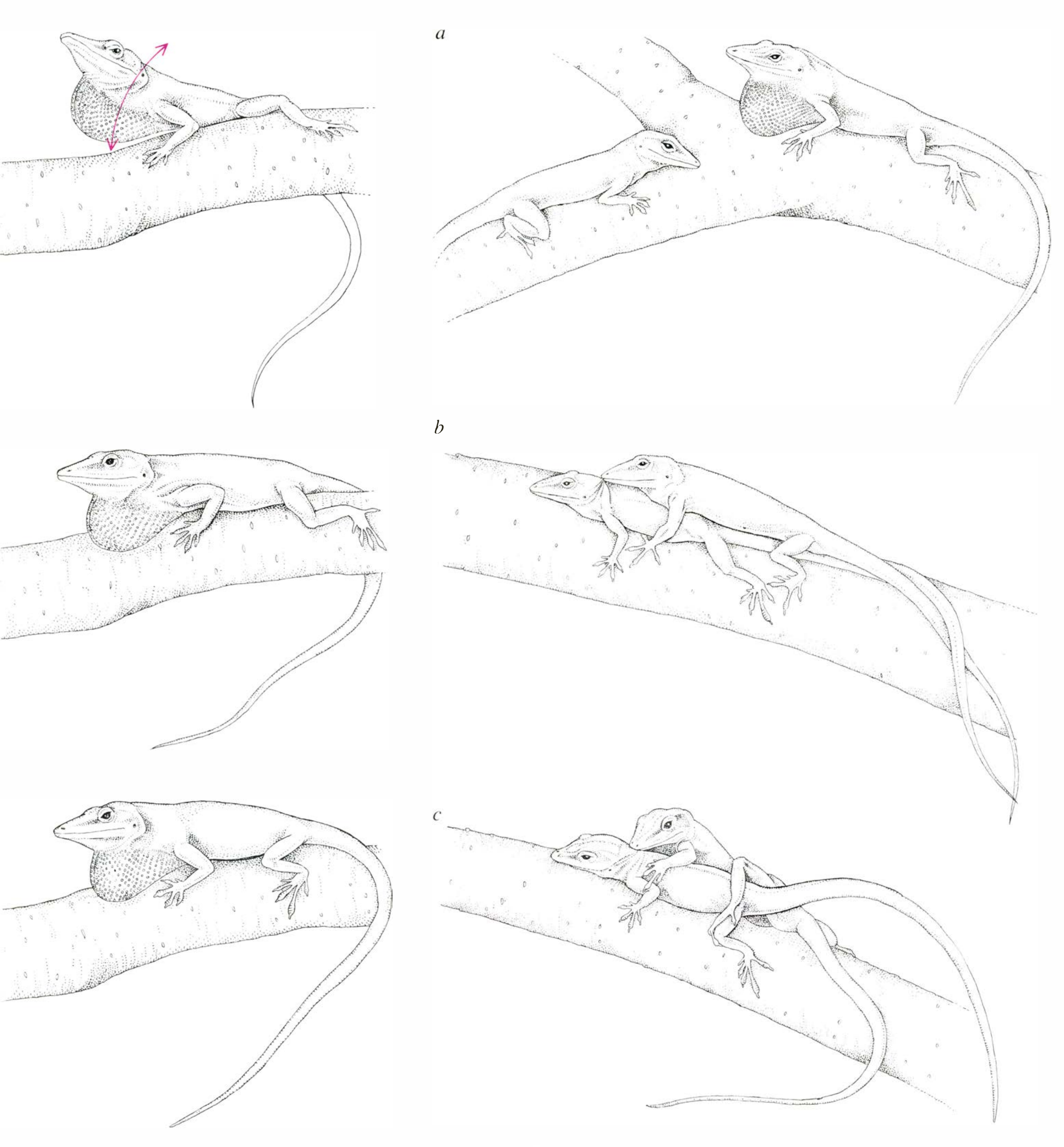


EXHIBICION DE DESAFIO que realiza un macho residente (*a*) cuando un macho extraño se introduce en su territorio; supone la extensión de la papada, una compresión lateral acentuada del cuerpo y una sacudida estereotipada del cuerpo. Si el intruso no responde con una exhibición de sumisión, el residente avanza hacia el intruso (*b*), frecuentemente con la boca abierta, y ambos animales luchan. Durante la lucha yerguen su cresta a lo largo del cuello y la región dorsal y aparece una mancha negra detrás del ojo; pueden trabar sus mandíbulas (*c*). El lance termina cuando un lagarto expulsa al otro.

DESAFIO Y SUMISION. La secuencia empieza (*a*) cuando un macho extraño, aquí el que hay a la izquierda, penetra en el territorio de un macho residente. El residente responde con una exhibición de desafío. Esta vez, sin embargo, el intruso no reacciona agresivamente, sino que ofrece una

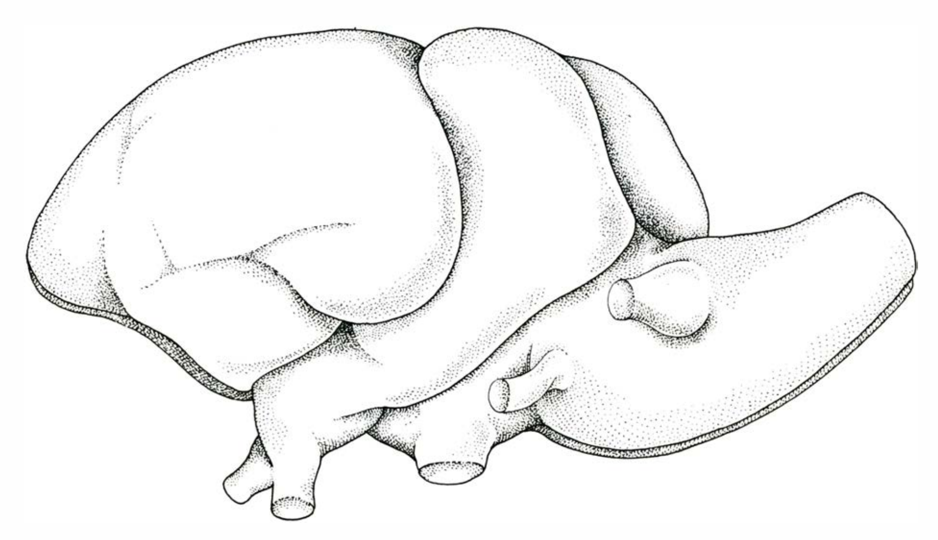
dos de tales órganos, cada uno de los cuales recibe el nombre de hemipene).

Se ha averiguado mucho sobre el control fisiológico del comportamiento reproductor del macho de *A. carolinensis*. Como era de suponer, el comportamiento sexual masculino depende de las gónadas. La castración conduce a una rápida disminución de la actividad de cortejo, pero la administración de andrógenos la reinstaura. La castración parece afectar mucho menos al comportamiento agresivo, que, sin embargo, depende no poco de los factores ambientales. Por ejemplo, si después de castrarlo devolvemos el macho a su propia jaula, su comportamiento agresivo se reduce paulatinamente o bien se mantiene, mientras que si se le coloca en una jaula que no conoce, la agresividad disminuye rápidamente.



exhibición de sumisión (*b*), que supone un rápido asentimiento con la cabeza, y cambia su color de verde (*claro*) a pardo (*oscuro*). Con esta respuesta, el macho residente no yergue su cresta, ni se forma la mancha negra inmediatamente detrás del ojo (*c*). Y el intruso acaba por marcharse.

EL CORTEJO Y EL APAREAMIENTO del anolis verde comienzan (*a*) cuando un macho sexualmente activo ve a una hembra y responde con una exhibición de cortejo, que incluye la extensión repetida de la papada (del modo que se indica en la fotografía de la página 95) y la serie de movimientos de asentimiento con la cabeza, que son característicos de cada individuo. Si la hembra es sexualmente receptiva, arquea su cuello y permite que el macho se acerque y lo agarre (*b*). Al montar a la hembra (*c*), evagina uno de sus dos hemipenes: el izquierdo si se halla sobre el lado derecho de la hembra y viceversa.



EL CEREBRO DEL ANOLIS VERDE tiene sólo el 0,003 por ciento del tamaño del cerebro humano; aquí se ilustra con un aumento de unos 20 diámetros. Debido a su pequeñez, se necesitó un aparato estereotáxico especial para implantar en él minúsculos tubos y agujas con fines experimentales.

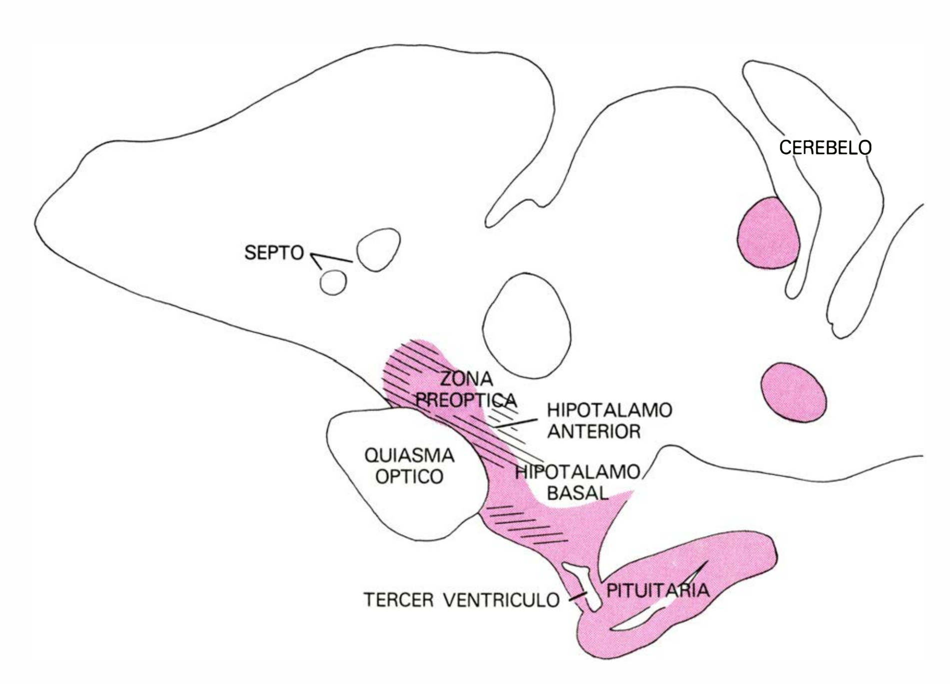
Se ha supuesto durante mucho tiempo que, en los animales vertebrados, el comportamiento reproductor del macho es disparado por la acción de los andrógenos sobre el cerebro. Recientemente se ha avanzado una hipótesis alternativa por parte de Frederick Naftolin, de la Facultad de Medicina de la Universidad de Yale. En síntesis, la hipótesis viene a afirmar que la hormona "masculina" testosterona se convierte en el cerebro, y por la acción de enzimas, en una hormona "femenina", el estradiol, que entonces activa el comportamiento sexual del macho. El proceso que caracteriza el cambio estructural en la hormona se conoce como aromatización. Sin embargo, a la hora de estimular estructuras sexuales secundarias se cree que la testosterona actúa de un modo directo o a través de su conversión en andrógenos no aromatizables, como la dihidrotestosterona.

Varios estudios apoyan la idea de que en el cerebro de muchas especies de mamíferos tiene lugar de forma normal la aromatización de andrógenos a estrógenos (término colectivo que se aplica a un grupo de hormonas sexuales femeninas), y que en muchas especies es el estradiol, y no la testosterona, el activador del comportamiento sexual masculino. Gloria Callard, Zoltan Petro y Kenneth Ryan, de la Facultad de Medicina de Harvard, han identificado enzimas aromatizantes en el cerebro de tortugas y serpientes. Han señalado que cuando se incuban partes específicas del cerebro de la tortuga dulceacuícola *Chrysemys picta* con andrógenos que han sido marcados con andrógenos radiactivos, posteriormente se encuentran en el tejido algunos estrógenos marcados radiactivamente. Este y otros descubrimientos llevaron a dichos autores a proponer que la aromatización de los andrógenos es una propiedad antigua, desde el punto de vista evolutivo, del sistema nervioso de los vertebrados.

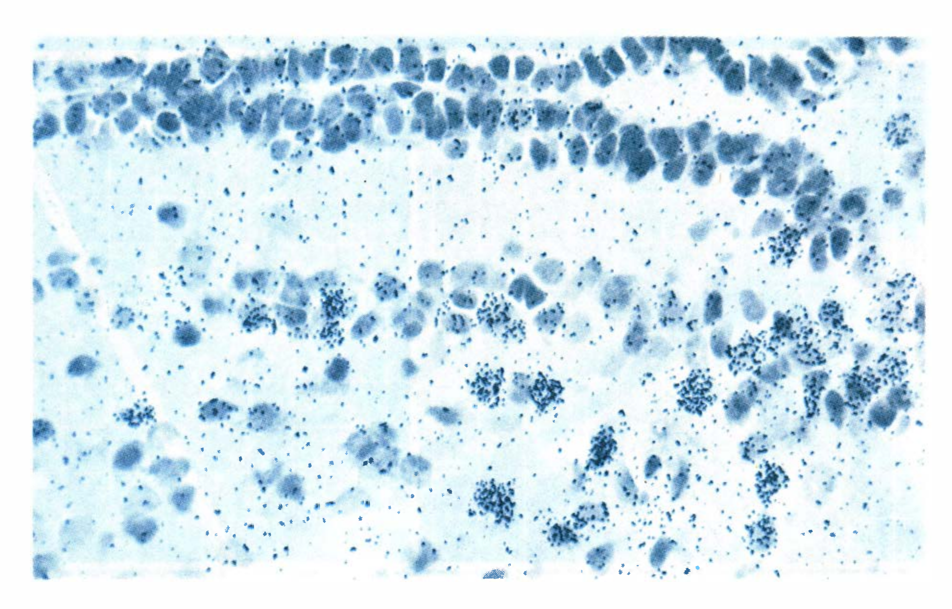
Antes de poder generalizar la hipótesis de la aromatización de los andrógenos a todos los reptiles, deben probarse en animales vivos los efectos de andrógenos aromatizables y no aromatizables. Los lagartos son muy adecuados para esta investigación. Tanto los saurios como las serpientes poseen una sección del riñón modificada para formar un segmento sexual. Esta estructura produce el líquido seminal y es semejante a la próstata y las vesículas seminales de los mamíferos. La actividad secretora del segmento sexual del riñón, que es muy sensible a la estimulación por andrógenos, puede determinarse midiendo el aumento de altura de las células epiteliales de los túbulos renales después de la estimulación hormonal.

Trabajando con esta medida como indicación de que las hormonas afectaban directamente a los órganos sexuales, y con la estimulación del comportamiento sexual tal como las exhibiciones de cortejo como indicación de que las hormonas actuaban sobre el cerebro, realizamos un experimento en el que se trataban machos castrados con testosterona, dihidrotestosterona o estradiol. El examen del segmento sexual del riñón reveló que sólo las dos primeras hormonas actuaban de forma directa sobre los órganos sexuales. Y únicamente la testosterona reinstauraba el comportamiento sexual.

Este hallazgo no excluye la posibilidad de que el estradiol actúe en el cerebro del macho. Varios investigadores han demostrado que la administración simultánea de dihidrotestosterona y estradiol estimula el comportamiento copulador en ratas castradas. Resultados de este tipo han conducido a la idea de que la actividad sexual masculina requiere, a la vez, la acción central o cere-



SECCION DEL CEREBRO mostrando las zonas (*color*) en las que se sitúan las células que concentran las hormonas esteroides sexuales, según revela la autorradiografía. En esta técnica, a lagartos machos y hembras a los que se les han extirpado las gónadas se les inyectan hormonas marcadas radiactivamente; después de sacrificar a los animales se colocan secciones de su cerebro sobre portaobjetos recubiertos de emulsión fotográfica. Cuando se revela esta emulsión, las zonas expuestas por la radiactividad indican dónde se han concentrado las hormonas. Las líneas oblicuas indican en qué lugares las lesiones cerebrales eliminan el comportamiento sexual y agresivo del macho, y las líneas horizontales aquéllos en los que las lesiones reducen la función pituitaria del lagarto anolis verdé.



CONCENTRACION DE ESTRADIOL en el cerebro de una hembra de *A. carolinensis*. Viene indicada por los granos negros en esta autorradiografía de una sección a través del núcleo periventricular del hipotálamo. El aumento es de 550 diámetros. El estradiol inyectado estaba marcado radiactivamente, de modo que expuso los granos de plata de la emulsión sobre la que se colocó la sección del cerebro.

bral del estradiol y la acción periférica de los andrógenos para estimular los órganos sexuales. En apoyo de esta hipótesis puede citarse el trabajo de dos grupos de investigadores, uno de ellos encabezado por Julian M. Davidson, de la Universidad de Stanford, y el otro por Paula Davis y Ronald J. Barfield, de la Universidad de Rutgers. Estos autores han observado que los implantes de estradiol en el cerebro de ratas únicamente inducen el comportamiento sexual si al mismo tiempo se administra dihidrotestosterona por vía sistémica.

Mis colegas y yo, en la Universidad de Harvard, examinamos también la posibilidad de que en el anolis verde la dihidrotestosterona y el estradiol actuaran al unísono. Cuando a lagartos castrados se les administraba simultáneamente ambas hormonas, la mitad de los animales (en dos experimentos separados) exhibieron súbitamente el modelo reproductor completo en un período de prueba único. Los resultados fueron muy distintos de los que se obtenían cuando sólo se administraba testosterona; esta hormona inducía típicamente una restauración gradual del comportamiento a lo largo de varios días. No cabía, pues, la menor duda de que el estrógeno podía desempeñar un papel tan vital en el control del comportamiento sexual masculino en los lagartos como lo hacía en algunos mamíferos.

Sin embargo, la otra mitad de los lagartos castrados en nuestros experimentos no mostraron ninguna alteración en el comportamiento sexual después del tratamiento combinado. Los animales destinados a nuestras experiencias fueron recolectados de diversos lugares, de modo que era posible que la variabilidad ambiental reflejara diferencias intrínsecas entre poblaciones en la cantidad de enzima aromatizante en el cerebro reptiliano. La variabilidad de comportamiento podía reflejar asimismo diferencias genéticas en la sensibilidad de un individuo dado a las hormonas.

La variabilidad individual en la respuesta le hace pensar a uno en la cordura que puede representar fiar en cepas de animales muy endogámicos para las pruebas de medicamentos. Muchas de las cepas endogámicas de roedores proceden de repetidas reproducciones consanguíneas entre hermanos y hermanas de forma que cada cepa se parece más a un individuo único y no refleja el espectro normal de variabilidad que se encuentra en la mayoría de poblaciones naturales. Este punto importante se suele pasar por alto en la interpretación de los datos sobre efectos de los medicamentos y en la determinación de los tratamientos clínicos apropiados.

Las hormonas regulan el comportamiento y la función de la glándula pituitaria a través de su influencia sobre el sistema nervioso central. Recientemente, nuestro grupo ha comenzado a estudiar la interacción de las hormonas, el cerebro y el comportamiento en el anolis verde.

Aunque se dispone de mucha información sobre las zonas del cerebro de los mamíferos que participan en el control del comportamiento y la fisiología de la reproducción, no estaba seguro de que tal información pudiera extrapolarse a los reptiles, de manera que me asocié con Joan I. Morrell y Donald W. Pfaff, de la Universidad Rockefeller. Primero identificamos los lugares que, en el cerebro de *A. carolinensis*, concentran hormonas sexuales esteroides. Nuestra técnica fue la autorradiografía. A lagartos machos y hembras a los que se les habían extraído las gónadas se les inyectaba una hormona marcada radiactivamente: estradiol, testosterona o dihidrotestosterona. Dos horas después, se sacrificaba a los animales, y el cerebro y la pituitaria se extraían rápidamente y se congelaban. Se cortaban luego los cerebros en secciones finas en un cuarto oscuro, y se colocaban las secciones en portaobjetos que habían sido revestidos con emulsión fotográfica. La mitad de estas preparaciones se guardaban en cajas opacas durante seis meses, la otra mitad durante nueve meses. Cuando se extraían los portaobjetos y se revelaba la emulsión, las zonas con un agrupamiento denso de granos negros sobre los núcleos celulares indicaban los lugares en los que se había concentrado la hormona marcada radiactivamente.

Los resultados de este estudio indican que en estos lagartos, como en todas las demás especies de vertebrados que se han investigado, las hormonas esteroides sexuales se concentran en zonas específicas del cerebro: la zona preóptica, el hipotálamo anterior y basal y el sistema límbico. Se ha descubierto que estas estructuras desempeñan un papel crucial en la reproducción en mamíferos y aves. Un examen de su función en los reptiles, que poseen un nivel de organización más primitivo, arrojaría luz sobre la evolución de los mecanismos que controlan la reproducción.

Puesto que el cerebro de un macho adulto de *A. carolinensis* es bastante pequeño, y representa solamente el 0,003 por cien del tamaño del cerebro humano, tuvimos que desarrollar técnicas especiales con el fin de investigar la función de las zonas en las que se concentraban los esteroides. Nuestros esfuerzos deben mucho al trabajo de Neil B. Greenberg, de la Universidad de Tennessee, que diseñó un aparato estereotáxico con el que se puede dirigir de manera precisa la punta de una aguja o de un tubo ultrafinos, en las tres dimensiones del espacio, hasta un punto determinado de un pequeño cerebro. Este autor preparó asimismo un atlas del cerebro anterior de *A. carolinensis*; trabajando con este atlas como si fuera un "mapa de carreteras" del cerebro, empezamos a estudiar el papel que las distintas zonas sensitivas a las hormonas desempeñan en el control del comportamiento y la fisiología de la reproducción del animal.

Una gran cantidad de datos de investigación en mamíferos y aves indica que

la región del hipotálamo anterior-zona preóptica constituye un centro importante para la integración del comportamiento. Puesto que a la vez es un lugar principal de captación de esteroides en el lagarto, examinamos primero su papel en la regulación del comportamiento reproductor del macho. La destrucción de esta zona en machos intactos y sexualmente activos causa la suspensión inmediata del comportamiento sexual, mientras que en machos castrados sexualmente inactivos la implantación directa de cantidades mínimas de testosterona en la zona preóptica reinstaura el comportamiento sexual.

Cuando se destruye la región inmediatamente anterior a la zona preóptica o el hipotálamo anterior basal, se corta la liberación de gonadotropina a partir de la pituitaria y los testículos degeneran pronto. En el primer caso, el efecto se da quizá sobre las células que producen las hormonas disparadoras; se han encontrado células de este tipo en el cerebro anterior de gran variedad de animales. La región donde estas hormonas pasan a la circulación pituitaria es el hipotálamo anterior basal.

Debe tenerse en cuenta que el comportamiento de un animal no es únicamente una consecuencia de la acción de las hormonas sobre zonas discretas del cerebro. Varios investigadores han demostrado que, en los mamíferos y las aves, la retroacción de las sensaciones periféricas que proceden de las estructuras sexuales secundarias incide mucho en la coordinación y consumación de las pautas de comportamiento. Esta relación es evidente en el comportamiento de cópula de los lagartos macho.

Podemos determinar cuál de los dos hemipenes está utilizando un macho de lagarto o de serpiente, observando la dirección hacia la que está vuelta la cola. Si el macho se halla sobre el lado izquierdo de la hembra, introducirá el hemipene derecho curvando su cola hacia la derecha, y viceversa. Los experimentos han demostrado que las sensaciones procedentes de los hemipenes desempeñan un papel importante no sólo en el control de la orientación inicial del macho durante la cópula, sino también en la conclusión de la misma. Este hallazgo concuerda con gran número de estudios en mamíferos que demuestran que la retroacción sensorial procedente del pene es necesaria para el comportamiento normal de apareamiento.

El comportamiento del macho que resulta de las interacciones hormona-cerebro tiene una influencia destacada en la fisiología de la reproducción de la hembra. Si se expone a hembras reproductivamente inactivas a un ambiente primaveral cuando se hallan solas o bien con otras hembras, se les estimula el crecimiento de los ovarios. Si exponemos una hembra de éstas a un macho sexualmente activo que la corteje, la tasa de desarrollo ovárico aumenta significativamente, lo que indica que el comportamiento del macho está facilitando los efectos estimuladores del ambiente. De hecho, se observa que no sólo el comportamiento de cortejo del macho es necesario para la secreción normal de gonadotropina por parte de la pituitaria de la hembra, sino que también la cantidad de gonadotropina secretada se halla relacionada con la cantidad o frecuencia del comportamiento de cortejo con que el macho asedia a la hembra.

Por otro lado, el comportamiento agresivo de los machos no sólo no facilita los efectos del ambiente cambiante, sino que los inhibe por completo. Las hembras alojadas con machos que continuamente se hallan luchando entre sí no presentan crecimiento ovárico, o bien éste es muy reducido. Por lo que sé, este descubrimiento es la primera demostración de que es posible "conectar" y "desconectar" la fertilidad de la hembra variando el ambiente del comportamiento social.

Durante la estación reproductora, la hembra desarrolla un folículo ovárico único, que es ovulado cada dos semanas aproximadamente. Como ocurre en muchas hembras humanas, los ovarios alternan en la producción del folículo, de modo que una hembra de anolis sexualmente activa tendrá típicamente en un ovario un folículo que se halla en un estado más avanzado que el mayor de los folículos en el otro ovario. Además, y como en muchos mamíferos, una hembra de *A. carolinensis* atraviesa ciclos de receptividad sexual que están relacionados con el modelo de crecimiento folicular. Presumiblemente, los ciclos reflejan fluctuaciones correspondientes de esteroides sexuales en la sangre.

Como cabría esperar, la extirpación de los ovarios anula la receptividad sexual en la hembra del anolis verde. La administración de estrógeno a una de estas hembras reinstaura el comportamiento de un modo relacionado con la dosis. La dosis umbral parece ser de 0,6 microgramos de benzoato de estradiol. Una única inyección de 0,8 microgramos induce la receptividad en el 85 por ciento de las hembras en 24 horas, y en todas ellas en 48 horas.

En muchos mamíferos la progesterona es importante en la regulación de la receptividad femenina. No funciona sola, pero cuando es administrada en un período apropiado después de haber "preparado" al animal con estrógeno, las dos hormonas actúan de forma sinérgica para facilitar la receptividad. Los estudios sobre el modelo de secreción hormonal que realizan los ovarios en estos animales revelan que esta facilitación que hace la progesterona coordina la receptividad de la hembra con la ovulación.

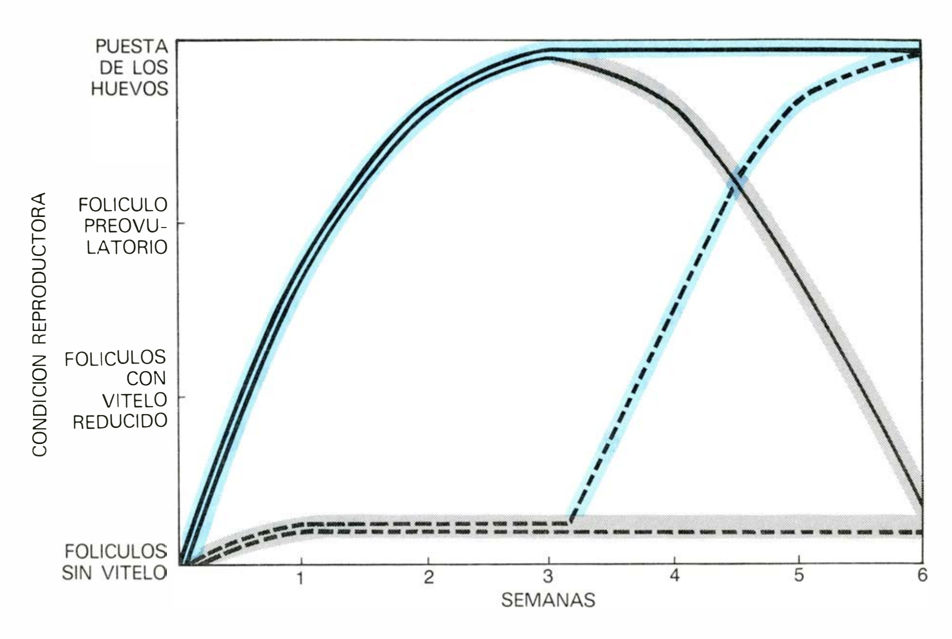
La progesterona parece realizar una función similar en los saurios. Por ejemplo, una hembra a la que se le han extirpado los ovarios no será receptiva al cortejo del macho después de una dosis subumbral de estradiol (0,4 microgramos), a menos que esta dosis sea seguida 24 horas más tarde por 60 microgramos de progesterona.

Como en los mamíferos, la progesterona puede tener un efecto muy distinto si se administra simultáneamente con

CICLO SEXUAL de la hembra del lagarto anolis verde, representado esquemáticamente. Las zonas blancas indican la época en la que la hembra no es sexualmente receptiva, el color claro simboliza el período de inicio de la receptividad y el color oscuro el período receptivo. Los ovarios se alternan en la producción del folículo; su estado de madurez se indica por el tamaño con que aparecen representados.

estrógeno o bien se administra después de un período suficientemente largo. En este caso inhibe el efecto estimulador del estrógeno. En el lagarto, una inyección única de 160 microgramos de progesterona 48 horas después de una dosis supraumbral (0,8 microgramos) de estradiol inhibe por completo el efecto del estrógeno.

Tanto en la hembra como en el macho, los estímulos centrales y los periféricos participan en la regulación de la receptividad. Cuando colocamos una hembra que es preovulatoria o que ha sido preparada con estrógeno en una jaula donde hay un macho, éste empieza a cortejarla de inmediato. Sin embargo, menos de un minuto después de la cópula la hembra deja de mostrarse receptiva al cortejo. Los experimentos han demostrado que el disparador de la transformación es la introducción del hemipene. Una hembra preovulatoria que es cortejada y montada pero no copulada continúa siendo receptiva. La duración de la introducción parece asimismo ser importante, puesto que las hembras cuya cópula se ha interrumpido continúan siendo receptivas hasta que se les permite consumar completamente la cópula con el macho.



INTERACCION del comportamiento del macho y del estado reproductor de la hembra en *A. carolinensis*. Se trabajó con cuatro grupos de hembras. Uno de ellos (*curva de abajo*) fue expuesto únicamente al comportamiento agresivo de los machos. Otro (*curva de arriba*) fue expuesto sólo a galanteo por parte de los machos. La curva que sube a las tres semanas corresponde a un grupo que estuvo expuesto a tres semanas de agresión y otras tres de galanteo; la curva que baja a las tres semanas representa a un grupo que recibió el tratamiento opuesto. Los períodos de cortejo están en color, los de agresión en gris. La condición folicular (y por tanto el estado reproductor de las hembras) se indica a la izquierda.

Hemos observado que la presencia de los ovarios es capital para la eliminación de la receptividad. Si hembras carentes de ovarios pero preparadas con estrógeno se aparean, vuelven a ser receptivas a las 24 horas. De hecho, los experimentos anteriores indican que estas hembras sólo dejan de ser receptivas unas seis horas después del apareamiento. Las hembras normales, en cambio, continúan frígidas durante varios días, lo cual sugiere que algún cambio en la producción de hormonas ováricas es la clave de la inhibición a largo plazo de la receptividad.

De la misma manera que el comportamiento del macho influye sobre la fisiología de la reproducción de la hembra, la presencia de hembras tiene un fuerte efecto sobre el macho. Los machos alojados con hembras presentan un desarrollo testicular y espermático más rápido que los machos encerrados con otros machos.

En resumen, la secuencia de acontecimientos en el ciclo reproductor del lagarto comienza con la estimulación ambiental de la actividad testicular. Las hormonas producidas por los testículos como resultado de esta estimulación actúan sobre zonas discretas del cerebro para influenciar la secreción de gonadotropina por parte de la glándula pituitaria y también para activar el comportamiento sexual del macho. Según el comportamiento del macho, se estimula o bien se inhibe la secreción de gonadotropina en la hembra, y por tanto también el crecimiento ovárico. El comportamiento sexual de la hembra, que depende en parte de hormonas ováricas y en parte de estímulos externos, influye después sobre el comportamiento y la actividad testicular del macho.

Los experimentos que he descrito indican que *A. carolinensis* es un animal adecuado para la investigación de fenómenos biomédicamente importantes que en los mamíferos son difíciles de aislar y separar en sus partes componentes, en particular en lo que se refiere a estudios de las patologías neurológicas que se hallan detrás de las disfunciones sexuales. Nuestros estudios han revelado asimismo cierto grado de conservadurismo en el control neural y endocrino del comportamiento reproductor en los vertebrados. Por ejemplo, hemos encontrado que las zonas del cerebro que concentran hormonas esteroides son similares en saurios y mamíferos, y hemos descubierto que la función de estas zonas en la regulación del comportamiento reproductor es asimismo parecida. También hemos hallado que la progesterona tiene un papel distintivo en el control hormonal de la receptividad sexual de la hembra (ya sea estimulándola o inhibiéndola) en los saurios al igual que en muchos mamíferos, lo que sugiere que este efecto bifásico de la hormona progesterona es antiguo en términos evolutivos.

Resulta asimismo digno de mención que las múltiples características peculiares de los lagartos y la integración de los descubrimientos de laboratorio con las observaciones en la naturaleza han conducido a una comprensión más clara del modo en que el comportamiento interactúa con la fisiología en la adaptación de un animal a su ambiente. Así, el estudio sugiere que la típica secuencia temporal observada en muchas especies entre la agresión y el cortejo, y los mecanismos fisiológicos subyacentes a estas acciones, pueden reflejar orígenes evolutivos comunes. Los efectos opuestos en las hembras de la agresión y el cortejo de los machos pueden explicar, a su vez, las pautas que se observan frecuentemente en los vertebrados de la zona templada, en los que los machos establecen territorios en las áreas reproductoras antes de la llegada de las hembras. Las hembras que llegan primero se verían impedidas para reproducirse debido al comportamiento agresivo que los machos demuestran entre sí, mientras que las hembras que llegaran más tarde serían rápidamente estimuladas por el comportamiento de cortejo de los machos territoriales. Finalmente, la conclusión de la receptividad sexual por medio del apareamiento tiene varias ventajas para la hembra: ésta se limita a elegir a un macho y disminuye su vulnerabilidad a los depredadores al reducir el tiempo que pasa apareándose, por lo que aumenta las probabilidades de sobrevivir para dejar un gran número de descendientes.

Atmósfera y contaminación atmosférica

La atmósfera no es pasiva frente a la contaminación; dispone de mecanismos para diluir los contaminantes que se inyectan en ella, pero esta capacidad de dilución es siempre limitada

Manuel Puigcerver

Sobre la contaminación atmosférica se habla y se escribe mucho desde hace años; a veces da la impresión de que demasiado. No es que la importancia de la cuestión se haya exagerado. Lo que sucede es que se ha puesto de moda, y con ello ha desbordado los límites de la comunidad científica –es decir, el ámbito donde los problemas se estudian de manera serena, objetiva e informada– para convertirse en tema de discusión pública y saltar de vez en cuando a la primera página de los periódicos. Como resultado, mucho de lo que se escribe adolece con frecuencia de insuficiente conocimiento por parte del autor o de opiniones particulares de éste que se presentan, intencionadamente o no, como hechos indiscutibles. Y ello es peligroso; porque si es cierto que cuando un problema preocupa realmente a la sociedad es quizá más fácil persuadirla a que aporte los medios económicos necesarios para abordarlo y tal vez resolverlo, no es menos cierto que la presión de una opinión pública confusa y desorientada puede obligar a los órganos de decisión a promulgar medidas improvisadas e imprudentes en lugar de hacerlo a partir de propuestas lógicas y razonadas elaboradas por expertos. Ahora bien, casi espontáneamente surge una pregunta obvia: ¿Quiénes son los expertos en contaminación atmosférica?

La respuesta es mucho menos obvia debido a la naturaleza interdisciplinaria del problema. Para darse cuenta basta con examinar algunos de sus elementos. Empecemos por definir la contaminación atmosférica como la impurificación de la atmósfera por inyección y permanencia temporal en ella de materias –gaseosas, líquidas o sólidas– ajenas a su composición normal o en proporción claramente superior a la de aquella. Estas substancias no son necesariamente tóxicas, nocivas o irritantes ni son siempre antropogénicas, es decir, originadas como consecuencia de la actividad del hombre. La erosión de la superficie terrestre por el viento, el transporte de polen, los incendios forestales o las erupciones volcánicas, que lanzan a la atmósfera cantidades ingentes de gases y cenizas, son ejemplos de contaminación atmosférica natural. Pero cuando se habla de contaminación atmosférica se piensa casi automáticamente en la producida por el hombre, y a ésta nos vamos a ceñir en el resto del artículo.

Entre los procesos que componen la contaminación artificial hay que señalar en primer término la emisión de contaminantes, que se lanzan a la atmósfera y permanecen en ella durante un tiempo más o menos largo ("tiempo de residencia"). Responsables de la emisión son las centrales térmicas que queman carbón o combustibles derivados del petróleo, las instalaciones industriales de diversas clases, los vehículos automóviles y otros medios de transporte (incluyendo ferrocarriles, barcos y aviones) y, finalmente, las actividades domésticas.

La segunda parte de la definición consideraba la permanencia de los contaminantes en la atmósfera durante cierto tiempo, que está relacionado con la capacidad de ésta para dispersarlos, con su reactividad química y con el lavado de la atmósfera que efectúa la lluvia, efectos todos que contribuyen a ir reduciendo la concentración de contaminantes.

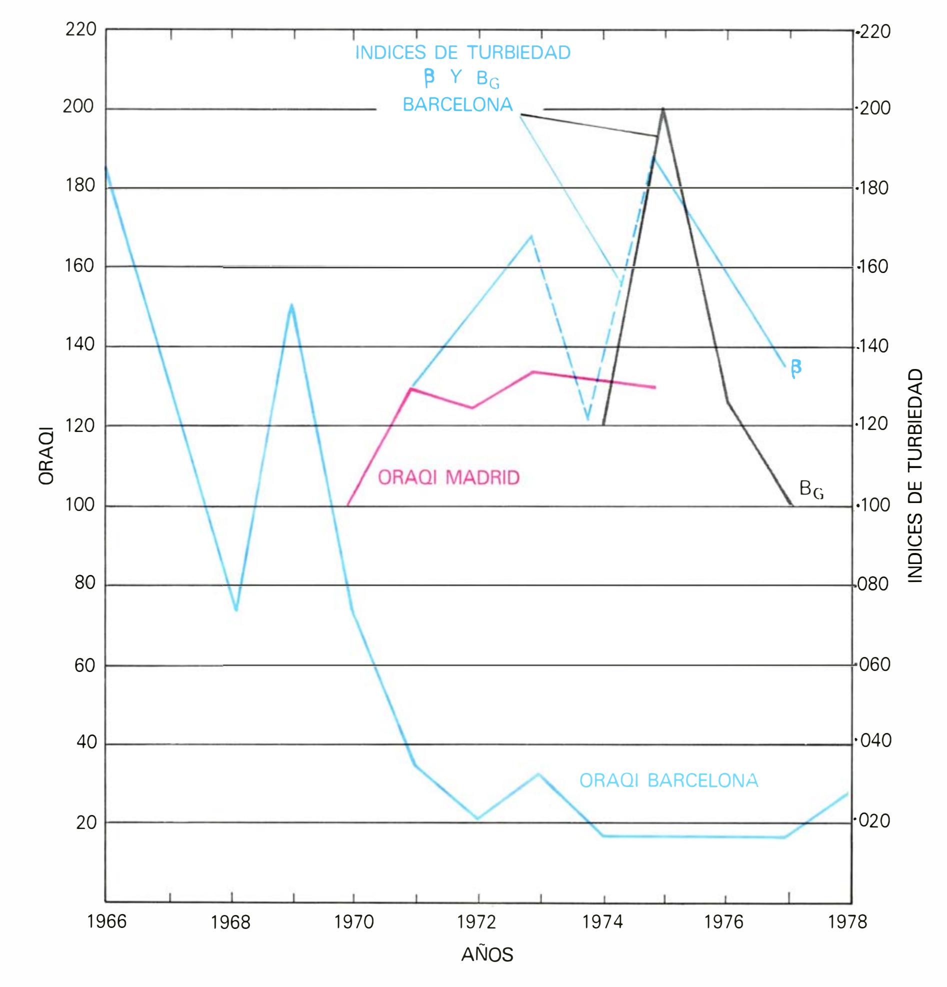
Finalmente, el estudio de la contaminación atmosférica ha de incluir el de los efectos de los contaminantes, tanto sobre seres vivientes como sobre estructuras (metálicas o de cemento, ladrillo, piedra, etc.), suelos, o sobre las propiedades de la atmósfera misma (reducción de la visibilidad, absorción y/o difusión de la radiación solar y terrestre, alteración del balance de calor del sistema tierra-atmósfera con las posibles influencias sobre el tiempo y el clima locales, etc.).

El examen de los tres puntos anteriores deja claro el carácter interdisciplinario del tema. En cuanto a la emisión, su regulación (imperativa en ciertos casos, conveniente siempre) lleva consigo: 1) la promulgación de una legislación adecuada, que obviamente compete a juristas conocedores del problema o bien asesorados; 2) el estudio de las características meteorológicas del lugar propuesto para nuevas instalaciones, que debería ser un prerrequisito obligado si la preocupación por la conservación del medio ambiente es real y verdadera y no, como tantas veces parece, un pretexto para demagogia y manifestaciones vocingleras; y 3) la solución de problemas típicamente tecnológicos, tales como la configuración, altura y tiro de las chimeneas industriales. Resulta, pues, evidente que todo intento de tratamiento global del problema necesita de la colaboración de un grupo numeroso de personas con muy distinta formación: abogados expertos en legislación, ingenieros, físicos, químicos, biólogos, médicos, veterinarios y finalmente, pero no en último lugar, meteorólogos. Como un proyecto de esta índole requiere la creación de organismos especializados, a la lista anterior habría que añadir personal técnico y administrativo de diversos grados.

Sin embargo, no hay que olvidar –y se olvida con harta frecuencia– a la atmósfera, que no se limita al papel de sujeto pasivo de la contaminación, sino que interviene activamente en el proceso desde el mismo momento en que los contaminantes le son inyectados. No se puede pensar en la contaminación atmosférica como en algo análogo a la inyección y ulterior difusión de un líquido colorante en un tanque de agua. La atmósfera es un sistema termo e

hidrodinámico de extraordinaria complicación, en el que coexisten movimientos de escalas tan diferentes como la hemisférica y la molecular junto con toda una gama de movimientos intermedios. La interacción y la redistribución constante de energía entre ellos es regla más bien que excepción, y una analogía sencilla como la difusión del colorante no sirve de modelo para el comportamiento de la atmósfera. Así va emergiendo el otro requisito que, además de la competencia en su propio campo de actividad, debe exigirse al hipotético grupo de personas calificadas de expertos en contaminación atmosférica. Una de las más distinguidas autoridades en la materia, el profesor Arthur C. Stern, de la Agencia de Protección Ambiental de los Estados Unidos (E.P.A.) lo ha expresado sin paliativos en esta conocida frase: "Nadie, independientemente de su formación profesional, debería trabajar en el campo de la contaminación atmosférica a menos que conozca con suficiente profundidad el comportamiento de la atmósfera; ésta es la caracteristica que diferencia la contaminación del aire de otros campos de las ciencias del medio ambiente...". Un estudio a fondo de la contaminación atmosférica es pues necesariamente interdisciplinario, pero el conocimiento de la atmósfera es condición *sine qua non* para no decir necedades en este resbaladizo terreno.

El estudio desde el punto de vista físico de la atmósfera y su comportamiento es el objeto de la Meteorología. Cuando se aplica a la contaminación atmosférica de un lugar (por ejemplo, una ciudad o un valle), tienen importancia las peculiaridades orográficas y meteorológicas locales (régimen de vientos, nubosidad, albedo del suelo, etc.), en tanto que cuando interesa, por ejemplo, la distribución global de dióxido de carbono (CO_2) lo importante es el carácter hemisférico y aun planetario de la circulación atmosféria. Cualquiera que sea el caso, el comportamiento de la atmósfera va a depender de los valores de un conjunto de variables (presión atmosférica, temperatura, humedad, radiación solar recibida y radiación terrestre emitida, las tres componentes de la velocidad del viento, etc., a distintas alturas), variables que escapan por completo a nuestro control. Cuando se trata de mantener la concentración local de contaminantes por debajo de valores prescritos, como es obligado donde existe legislación al respecto, se puede regular la cantidad de contaminantes emitidos, eliminar total o parcialmente algunos de ellos, y variar el caudal (o masa por unidad de tiempo) eyectado; todo ello mediante una tecnología a veces muy avanzada y a un costo con frecuencia considerable. Por el contrario, *al hombre no le es posible alterar substancial y voluntariamente las condiciones meteorológicas.* (La alteración local del tiempo durante cortos intervalos –por ejemplo, la disipación de nieblas en aeropuertos mediante quemadores de petróleo– es prohibitiva salvo en situaciones límite, y no puede considerarse seriamente como una modificación real del tiempo por el hombre.) De aquí la importancia decisiva de la Meteorología en relación con la contaminación atmosférica, pese a que esta cuestión ni siquiera se menciona en gran parte de la abundante literatura sobre el tema. El papel preponderante que la Meteorología



EVOLUCION DE VARIOS INDICADORES del grado de contaminación de la atmósfera al transcurrir el tiempo. El ORAQI (*O*ak *R*idge *A*ir *Q*uality *I*ndex) es un índice de calidad del aire que se calcula a partir de las mediciones de los cinco contaminantes principales (oxidantes, óxidos de azufre, dióxido de nitrógeno, monóxido de carbono y partículas) y de las concentraciones de esos contaminantes que se consideran las máximas permisibles. Si éstas son precisamente iguales a los valores medidos, el índice Oraqi valdría 100; en el caso de aire muy puro vale 10. La curva roja ha sido obtenida por el profesor Catalá de Alemany y sus colaboradores E. Hernández y J. Hernández, de la Universidad Complutense de Madrid; la curva azul ha sido obtenida en Barcelona por el autor y sus colaboradores J. Lorente y S. Alonso. En uno y otro caso se ha hecho uso de los datos publicados por los respectivos servicios municipales de vigilancia. Los valores de la curva de Madrid son excesivamente altos hasta 1975. En cambio, la curva de Barcelona da a partir de 1970 valores completamente en desacuerdo con la realidad; en particular, a partir de 1974 la curva da una impresión bucólica muy lejos de aquella. Aceptando como buenos los valores de la red municipal de vigilancia, la discrepancia puede estribar en la manera de calcular el índice. Cuando no se tienen mediciones de los cinco contaminantes indicados, como es nuestro caso, la contribución de los demás se estima a partir de los valores medios de varias ciudades de los Estados Unidos en que se miden habitualmente los cinco contaminantes. Si alguno de los que no se miden en Barcelona tuviera normalmente valores muy superiores al promedio, la estima del índice quedaría desvirtuada. De todas formas es extraordinariamente sorprendente encontrar valores Oraqi inferiores a 20 en una ciudad de más de un millón de habitantes. En la figura se hace una representación análoga para ciertos índices de turbiedad obtenidos en Barcelona por Jerónimo Lorente y el autor. Los índices de turbiedad dan idea de la atenuación que los aerosoles producen en la radiación solar al atravesar la atmósfera. La diferencia principal entre β (coeficiente de turbiedad de Angström)y B_G (coeficiente de turbiedad de Schüepp-Volz) estriba en que en el último se realiza la medida precisamente a la longitud de onda de 0,5 micrometros. Cuanto mayores los coeficientes, mayor es la atenuación debida a aerosoles y por tanto mayor la concentración de éstos. Lorente ha mostrado que la concentración de aerosoles en microgramos por metro cúbico se puede estimar multiplicando B_G por 650. Las curvas muestran ciertamente una disminución del contenido de aerosoles con el tiempo, pero el contenido de aerosoles queda todavía muy por encima del valor rural que correspondería al oraqi.

desempeña en problemas relacionados con la contaminación atmosférica aparecerá una y otra vez en el resto de este artículo.

Antes de entrar en materia procede decir algo sobre los contaminantes usuales y la manera de medirlos. Una primera subdivisión distinguiría entre los contaminantes gaseosos y los que no lo son. Estos últimos incluyen partículas sólidas o líquidas de tamaño microscópico o submicroscópico pero superior al molecular. Se cuentan entre ellos el polvo, producido por desintegración mecánica natural o artificial de materia sólida, dando partículas de diámetro próximo al micrometro (un micrometro es la millonésima parte del metro o la milésima de milímetro, unidad antiguamente llamada micra); los humos, constituidos por partículas sólidas y líquidas formadas mediante procesos complicados (condensación de vapores sobresaturados, sublimación y reacciones químicas) subsiguientes a las combustiones y cuyo dominio de tamaños es vecino al del polvo; las brumas, constituidas por gotículas en suspensión, de unos 10 micrometros de diámetro, formadas por condensación de vapor sobre ciertas partículas higroscópicas llamadas núcleos de condensación o, más raramente, por pulverización de agua en la atmósfera; y finalmente el aerosol, término algo vago que incluye cualquier nube de partículas sólidas o líquidas, microscópicas o submicroscópicas de diversa naturaleza: humos, bruma, niebla, núcleos de condensación, etc. Estos impurificantes se suelen llamar "partículas" a falta de término mejor (los anglosajones han acuñado la expresiva palabra "particulates" a partir de "particulate matter").

Los contaminantes que llamamos gaseosos son aquellos que entran en la atmósfera como tales; algunos de ellos se asocian a gotitas de agua o núcleos higroscópicos y/o sufren reacciones químicas que conducen a productos insolubles, terminando su vida como partículas. Hay algunos, los llamados contaminantes secundarios, que no son directamente inyectados a la atmósfera, sino que se producen por reacción química entre los que sí lo son (primarios) y los gases atmosféricos: el ejemplo más típico es el ozono que aparece como consecuencia de complicados procesos fotoquímicos (es decir, inducidos por la luz solar) entre óxidos de nitrógeno e hidrocarburos. Entre los primarios, los más importantes son: *a)* Los compuestos de azufre, que generalmente penetran en la atmósfera en forma de dióxido de azufre o anhídrido sulfuroso SO_2, que puede oxidarse a SO_3 y éste, por hidratación, da lugar a ácido sulfúrico H_2SO_4. Este puede producir diversos sulfatos, designados genéricamente MSO_4. Igualmente pueden coexistir mezclas de diversos óxidos de azufre SO_x; y finalmente hay que mencionar el sulfuro de hidrógeno H_2S. *b)* Compuestos de nitrógeno, que generalmente se originan como óxido de nitrógeno NO o amoníaco NH_3 y mediante ulteriores transformaciones forman NO_2, mezclas de distintos óxidos de nitrógeno NO_x y nitratos o mezclas de ellos MNO_3; *c)* Compuestos de carbono, fundamentalmente el óxido de carbono CO que puede pasar por oxidación a dióxido de carbono CO_2; éste es un componente natural de la atmósfera [*véase la ilustración de esta página*] y no debe considerarse como contaminante excepto cuando su concentración excede claramente su valor normal. Hay que citar además los hidrocarburos alifáticos y aromáticos, aldehídos, cetonas y ácidos orgánicos, procedentes de la combustión incompleta de petróleos y gasolinas, del refinado de crudos o de los pulverizadores a presión. *d)* Finalmente, deben citarse por su carácter peligroso los halógenos HCl y HF producidos en ciertos procesos metalúrgicos, aunque en general su concentración es muy pequeña.

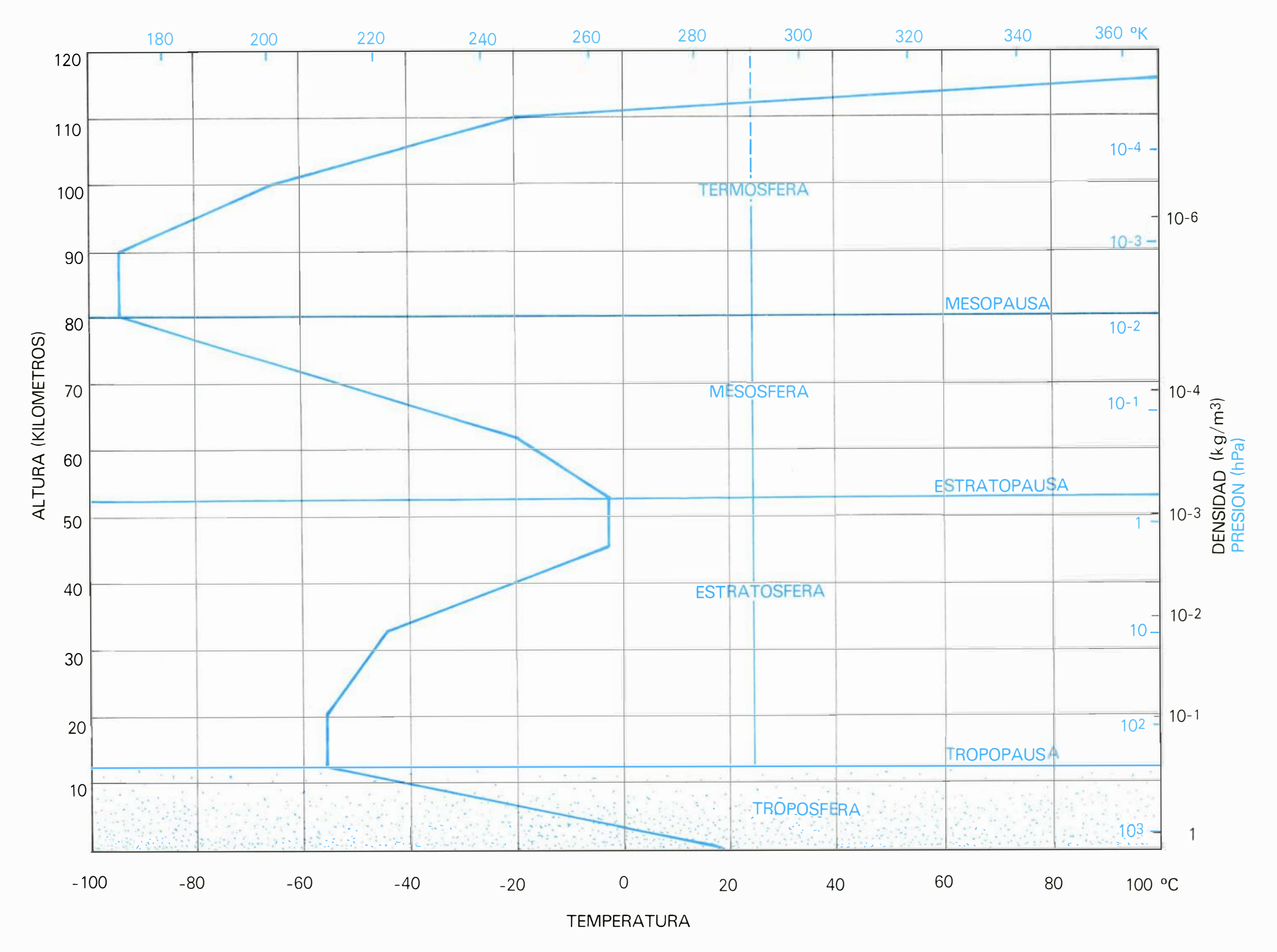
Para poder juzgar lo que se entiende por "muy pequeña" hay que indicar algo sobre el valor de las concentraciones de contaminantes. En el caso de gases, éstas se miden generalmente mediante captadores en los que la muestra de aire ambiente se hace pasar primero a través de un filtro (donde se depositan las partículas, cuya concentración se estima por métodos físicos) y luego a través de soluciones adecuadas para detectar químicamente los contaminantes de interés. En el caso de aerosoles, su concentración también se puede estimar indirectamente a partir del cambio que su presencia produce en las propiedades de la atmósfera, en particular en la atenuación de la radiación solar al atravesarla [*véase la ilustración de la página 105*]. Hablando en términos corrientes, las concentraciones son siempre muy pequeñas en valor absoluto, tanto que para expresarlas se ha tenido que recurrir a unidades desacostumbradas. La concentración de partículas se suele expresar en microgramos (es decir, millonésimas de gramo) de contaminante por metro cúbico de aire. En principio esta unidad también se podría usar para los contaminantes gaseosos, pero los químicos, que son quienes realizan los análisis, suelen preferir expresar concentraciones de gases en tanto por cien en volumen. Un valor típico de concentración de un contaminante gaseoso sería 0,0001 por ciento, cifra ciertamente poco cómoda. Por ello se acostumbra a multiplicar por 10.000, con lo que se obtiene la concentración en centímetros cúbicos de gas contaminante por metro cúbico de aire, o, como se designa habitualmente, "partes por millón" (ppm); el valor 0,0001 por ciento corresponde a una parte por millón.

Con lo anterior todavía no hay bastante para juzgar si la concentración ambiental de un contaminante (llamada a veces "inmisión", término evidentemente desafortunado) es alta o baja. Para ello es necesario, además, conocer los valores "de fondo" (es decir, los que se medirían en una atmósfera básicamente limpia como puede ser un medio rural alejado de ciudades e industrias) y los valores medios en el lugar considerado. Pero dado que no todos los contaminantes producen el mismo efecto sobre los seres vivientes, tampoco el valor de las concentraciones (ni aun con las referencias arriba indicadas) acaba de dar idea,

COMPOSICION DE LA ATMOSFERA SECA

	PARTES POR MILLON (VOLUMEN)	$\mu g/m^3$
NITROGENO	780.400	$8,95 \times 10^8$
OXIGENO	209.400	$2,74 \times 10^8$
AGUA	–	–
ARGON	9.300	$1,52 \times 10^7$
DIOXIDO DE CARBONO	315	$5,67 \times 10^5$
NEON	18	$1,49 \times 10^4$
HELIO	5,2	$8,50 \times 10^2$
METANO	1,0 - 1,2	6,56 $1,81 \times 10^2$
KRIPTON	1,0	$3,42 \times 10^2$
OXIDO NITROSO	0,5	$9,00 \times 10^2$
HIDROGENO	0,5	$4,13 \times 10$
XENON	0,08	$4,29 \times 10^2$

EN LA COLUMNA de la izquierda de esta tabla se da la cantidad de cada componente en centímetros cúbicos por metro cúbico de aire o "partes por millón" (ppm). En la de la derecha se da la masa de componente en microgramos (o millonésimas de gramo) por cada metro cúbico de aire.



ESTRUCTURA VERTICAL MEDIA de la atmósfera. El curso de la temperatura con la altura se representa mediante la curva de trazo grueso. Los valores de la presión y la densidad se indican a ciertas alturas. La presión se expresa en hectopascal (hPa), unidad que hasta época reciente se denominaba milibar; la presión atmosférica normal en el suelo es de 1013 hectopascal. La densidad está expresada en kilogramos por metro cúbico, y su valor en la vecindad del suelo es próximo a un kilogramo por metro cúbico. Se indican los nombres de las distintas capas atmosféricas y algunos detalles o fenómenos llamativos. Respecto al problema tratado en este artículo, la capa atmosférica de interés primordial es la más baja, llamada troposfera (*región sombreada*). En la troposfera, la temperatura disminuye por término medio 0,6 grados Celsius por cada 100 metros que aumente la altura, pero en sondeos atmosféricos individuales el valor es variable. La troposfera es la capa afectada por el grueso de la contaminación, y tanto más cuanto más baja la región considerada. La estratosfera, capa inmediatamente encima de la troposfera, se caracteriza porque en ella la temperatura se mantiene constante en su parte más baja para aumentar después cuando lo hace la altura. La capa que separa troposfera y estratosfera se llama tropopausa. Los contaminantes originados en el suelo raramente sobrepasan la tropopausa; ciertas grandes erupciones volcánicas han inyectado en la estratosfera cenizas y polvo, que han permanecido allí durante años. Los aviones reactores contaminan la estratosfera esporádicamente y en pequeña cuantía. El efecto que una contaminación de la estratosfera en mayor escala produciría sobre el clima de la Tierra no se conoce en forma inequívoca.

salvo al experto, de lo puro o deteriorado que está en el aire que respiramos. A este fin se han ideado diversos índices de calidad del aire que tratan de expresar mediante un número el efecto conjunto de los contaminantes más importantes. El mejor camino para formular tales índices consiste en comenzar por establecer un valor tipo para cada contaminante, valor que suele ser su máxima concentración permisible sin que se aprecien efectos adversos. Con ello, los valores tipo de los diversos contaminantes son aproximadamente equivalentes en cuanto a sus efectos nocivos o molestos. (No se puede omitir la palabra "aproximadamente" porque, siendo los efectos de los distintos contaminantes muy diferentes, decidir cuándo son equivalentes no es cosa fácil ni enteramente objetiva; por ello los valores tipo en distintos países difieren, a veces considerablemente.) Partiendo de estas ideas, W. A. Thomas y L. R. Babcock, entonces en el Laboratorio Nacional de Oak Ridge en Tennessee, U.S.A., propusieron en 1971 un índice que ha obtenido gran aceptación y quizá llegue a generalizarse, al menos en los países occidentales. Conocido por ORAQI (*O*ak *R*idge *A*ir *Q*uality *I*ndex), está basado en los cinco contaminantes (óxidos de azufre, de nitrógeno, de carbono, oxidantes y partículas) considerados fundamentales por la Agencia de Protección Ambiental del Gobierno de los Estados Unidos (Environmental Protection Agency, EPA), para los que ésta había establecido valores tipo. La concentración medida para cada contaminante se divide por el valor tipo y se suman las cinco fracciones así obtenidas. La suma se multiplica por un coeficiente y el resultado se eleva a un cierto exponente, elegidos ambos de manera que ORAQI = 10 cuando todos los contaminantes tengan el valor de fondo (aire limpio) y ORAQI = 100 cuando todos tengan el valor tipo (aire contaminado al máximo permisible). Un ORAQI mayor que 100 correspondería, según el criterio de E.P.A., a aire contaminado por encima del límite permisible. La ilustración de la página 105 permite comparar los valores del índice obtenidos a partir de los datos de las respectivas redes municipales de vigilancia por el profesor Catalá de

Alemany y sus colaboradores para Madrid (*color azul*) y por el autor y los suyos para Barcelona (*color rojo*).

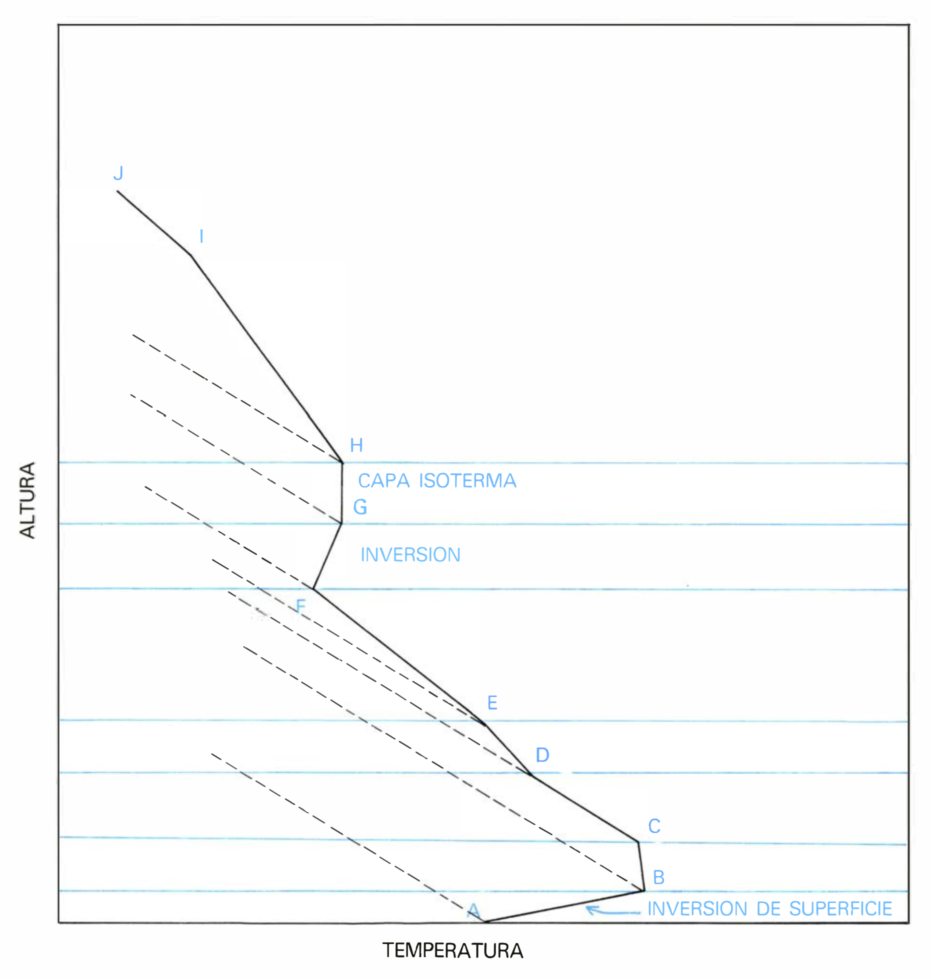
Se dijo más arriba que las condiciones meteorológicas juegan un papel primordial en problemas relacionados con la contaminación atmosférica. En efecto, tales condiciones, referidas a un período de tiempo corto (desde horas a pocos días), determinan el tiempo atmosférico. Referidas a los valores medios durante un período relativamente largo (más de treinta años) determinan el clima de un lugar. La distinción es importante porque vamos a necesitar hacer referencia a uno y otro y porque la ignorancia y la pedantería están en nuestro país fomentando el confusionismo al hacer un uso impropio de la palabra clima: recuérdese la frecuencia con que se lee u oye decir que un partido de fútbol o la salida de un vuelo fueron dificultados "por las condiciones climatológicas", cuando lo correcto –y más sencillo– sería decir "por el mal tiempo". El término "clima" no se puede aplicar a algo que sólo dura horas.

Del tiempo –es decir, de las condiciones meteorológicas en un día dado o durante una fracción de él– depende en último término la capacidad de la atmósfera para diluir hasta concentraciones aceptables los contaminantes lanzados a ella, y a su vez esta aptitud debería constituir el criterio limitante de las posibles emisiones. Del clima de un lugar o región depende el que las condiciones que favorecen la rápida difusión de contaminantes se den con mayor o menor frecuencia e intensidad o incluso que no se den. El conocimiento del clima y en particular de las condiciones en la capa atmosférica inmediata al suelo (capa fronteriza o capa límite) debería ser decisivo al proyectar la localización de industrias contaminantes.

En definitiva, el grado de contaminación atmosférica, es decir, la concentración de contaminantes en la atmósfera, depende por un lado de la cantidad de ellos lanzada a la atmósfera y por otro de la velocidad con que ésta los dispersa, la cual es función de ciertas condiciones meteorológicas que vamos a examinar. Se ve casi intuitivamente que si la atmósfera está en reposo, los contaminantes emitidos en su seno van a difundirse lentamente, de manera parecida al ejemplo del colorante en un tanque de agua. Cuanto más activo sea el movimiento vertical del aire y más veloz el movimiento horizontal, tanto más rápida será la difusión de contaminantes. En la atmósfera, el movimiento vertical en la escala de tamaños que afecta a la contaminación depende del carácter de la estratificación atmosférica, y el movimiento horizontal depende de la velocidad y rafagosidad del viento. Veamos por separado el papel de cada una de ellas.

El carácter de la estratificación atmosférica está relacionado con la variación de la temperatura a lo largo de la vertical. En la troposfera o parte baja de la atmósfera [*véase la ilustración de la página 107*], que es la de mayor interés a nuestros efectos, se encuentra que por término medio la temperatura desciende al aumentar la altura en una cuantía aproximada de 0,6 grados Kelvin por cada 100 metros de ascenso. Esta disminución se llama gradiente vertical medio de temperatura o gradiente geométrico de temperatura y se suele designar por α; se elige para definirlo la disminución de temperatura a fin de que resulte positivo si, como ocurre generalmente en la troposfera, la temperatura decrece cuando la altura crece.

Las cosas suceden de manera muy distinta cuando en lugar de valores medios se representan, por ejemplo, los correspondientes a un día dado. Estos se obtienen usualmente mediante un instrumento llamado radiosonda, dotado de sensores de presión, temperatura y humedad y de un pequeño radiotransmisor. El instrumento se eleva colgado



VARIACION DE LA TEMPERATURA con la altura en la troposfera en un día determinado; dista mucho de ser constante e igual a 0,6 grados Celsius por 100 metros. La figura representa una de tales curvas de variación, que muestra un carácter fuertemente irregular. Por encima del punto *C* hasta el *F* y por encima del punto *H*, la temperatura disminuye al aumentar la altura, pero no lo hace en forma constante, como lo demuestran los cambios de pendiente de la curva. Los segmentos de trazos representan porciones de las curvas de evolución de la figura siguiente, representados aquí a fin de poder comparar las pendientes. Hay una región (*GH*), llamada capa isoterma, en que la temperatura no varía con la altura, e incluso hay otras, como la *AB* y la *FG*, en que la temperatura, lejos de decrecer al aumentar la altura, aumenta con ella. Como en estas regiones se invierte la variación normal de la temperatura con la altura en la tropopausa, se llaman por ello inversiones. Una gráfica como la aquí representada se llama curva de estado porque es una especie de diagnóstico del estado de la atmósfera en ese momento. Tales curvas se obtienen corrientemente mediante instrumentos llamados radiosondas, dotados de sensores de presión, temperatura y humedad, además de un pequeño radiotransmisor. El radiosonda se eleva en la atmósfera colgado de un globo lleno de hidrógeno y el transmisor va comunicando a la estación receptora en tierra los valores de las tres variables indicados en cada momento por los sensores. Mediante un radioteodolito o un radar (en cuyo caso se añade al radiosonda una pantalla reflectora) se puede también determinar la altura y distancia del globo, lo que permite determinar el viento a distintas alturas. La curva de estado obtenida a una hora determinada puede diferir substancialmente de la que se encuentra horas antes o después, especialmente en su parte más baja; esta variabilidad es fundamental en relación con el tiempo y con la distribución de contaminantes.

de un globo lleno de hidrógeno, y durante el ascenso, la emisora va transmitiendo a la estación receptora en tierra los valores que sucesivamente va tomando cada variable. Hay además un sistema radiotelemétrico que permite determinar la posición del globo en cada instante y, a partir de ella, la altura y la dirección y velocidad del viento. Los datos suministrados por el radiosonda permiten construir una gráfica llamada curva de estado [*véase la ilustración de la página 108*]. En ella, en lugar de una disminución continua de la temperatura al aumentar la altura, se encuentra una disminución irregular: la pendiente varía de unas regiones a otras de la gráfica, existiendo zonas como la CD en que la temperatura disminuye más de 0,6 K por 100 metros, otras como las ED y HI en que lo hace menos y otras, como la GH, en que la temperatura no varía con la altura (capas isotermas). Finalmente, se encuentran a veces regiones en la curva de estado en las que la temperatura no sólo no disminuye al aumentar la altura, sino que aumenta con ella (capas AB y FG en la ilustración). Estas regiones se llaman inversiones porque en ellas se invierte el curso normal de la temperatura con la altura. En resumen, en las curvas de estado individuales el gradiente vertical de temperatura α no es constante ni vale 0,6 K/100 m, sino que varía de unas regiones a otras de la curva e incluso se hace nulo en las capas isotermas y negativo en las inversiones.

Quizá parezca sorprendente que la temperatura pueda aumentar en la atmósfera al aumentar la altura. Por ello y por la extraordinaria importancia de las inversiones en relación con la contaminación atmosférica, diremos más adelante unas palabras sobre algunos de los mecanismos que dan lugar a esta curiosa característica del perfil vertical de temperatura.

Las curvas de estado descritas arriba representan, como su nombre indica, la sucesión con la altura de los estados simultáneos de una columna de aire ambiente; son variables de un lugar a otro y, en el mismo lugar, de una hora a otra. Por supuesto, en el mismo lugar y a la misma hora, la curva de estado tiene en general distinta pendiente de una altura a otra. Así pues, las regiones singulares en las curvas de estado deben concebirse como algo cambiante, que se encuentra allí ahora pero que mañana o dentro de unas horas no estará a la misma altura ni tendrá la misma pendiente y quizá no aparecerá en absoluto. Debe evitarse pensar en las inversiones u otras características de la estructura atmosférica como algo permanente, algo que siempre se encuentra allí, creencia ésta que no por implícita está menos generalizada entre quienes conocen poco la atmósfera.

En estas condiciones parece difícil extraer de este caos algún orden para ver cómo se comportará la atmósfera respecto a un aglomerado de moléculas gaseosas que, como ocurre con un penacho, tiende a ascender en su seno. Hay sin embargo una manera sencilla de establecer criterios de comportamiento esencialmente ligados al carácter de la estratificación atmosférica.

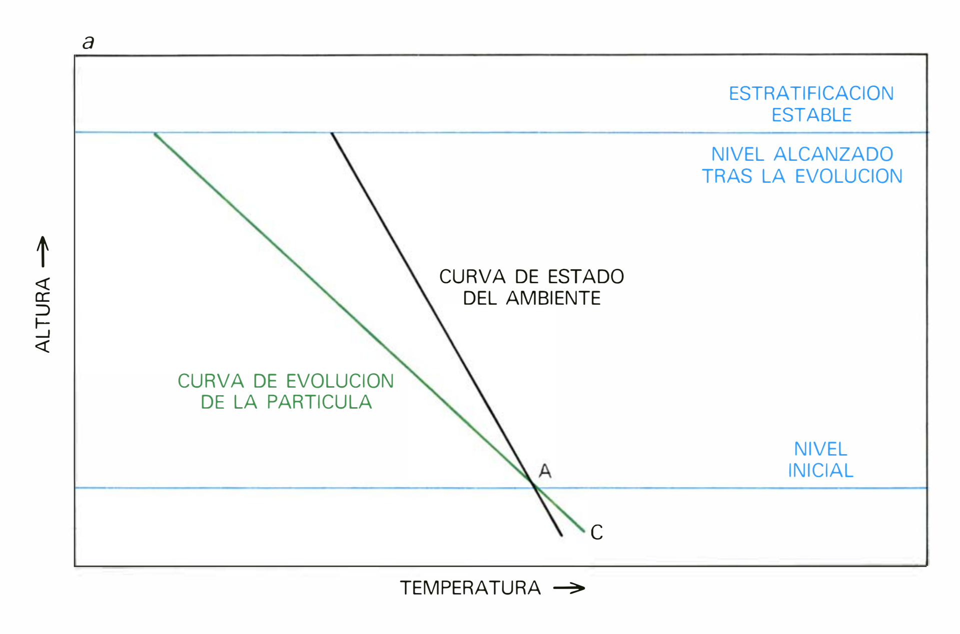
Una burbuja de aire que conserve su identidad al moverse en el seno de la atmósfera, de tal manera que siempre esté formada por las mismas moléculas gaseosas, constituye lo que en el lenguaje de la Mecánica de fluidos se acostumbra a llamar "partícula gaseosa" o simplemente partícula. Para evitar confusiones hay que prescindir aquí de la noción de partícula que se usa en Física Atómica o en Mecánica Cuántica: lo característico de la partícula, tal como se concibe en Mecánica de fluidos, no es el tamaño, sino la preservación de la identidad. De hecho, una partícula de aire como las que nos servirán en lo sucesivo para razonar puede tener un volumen de varios metros cúbicos.

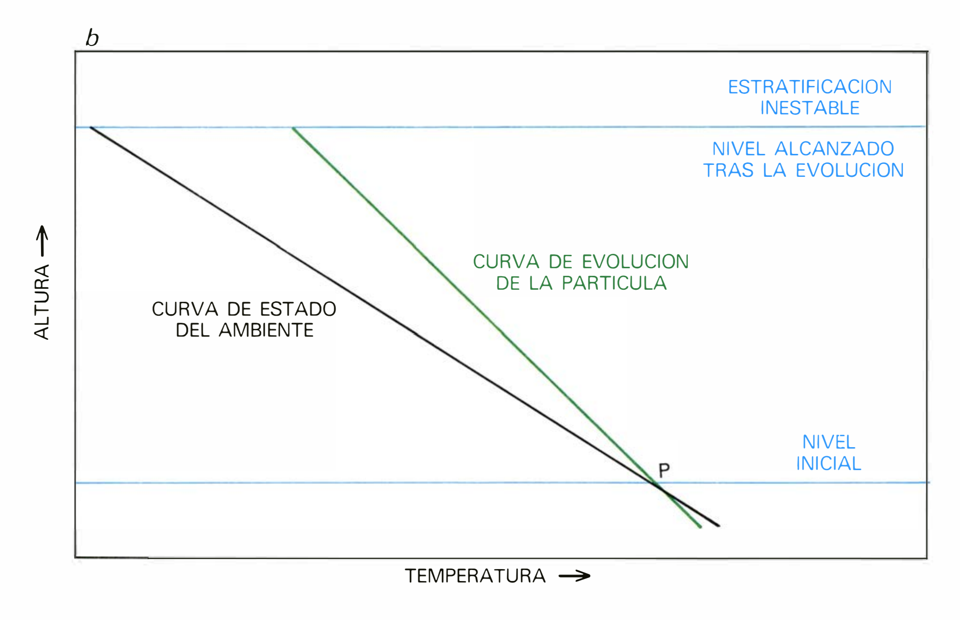
Supongamos que una de tales partículas de aire se eleva, por alguna razón, en el seno del aire ambiente. La razón puede ser un obstáculo orográfico, o la turbulencia atmosférica (de la que hablaremos más adelante) o el calentamiento desigual del suelo sobre el que reposa (por ejemplo un erial rodeado de sembrados o un calvero del bosque). Cualquiera que sea la causa, admitamos que la partícula se eleva respecto a su nivel inicial. Al subir va encontrando capas de aire ambiente cuya presión es cada vez más pequeña. Para ponerse en equilibrio de presión con ellas, la partícula se expande, mas para ello ha de efectuar cierto trabajo de dilatación. ¿De dónde saca la partícula la energía necesaria para efectuar este trabajo? Evidentemente, no de la radiación solar que la atraviesa, para la que el aire es prácticamente transparente y por tanto incapaz de calentarse absorbiéndola. Tampoco a la partícula le es fácil intercambiar calor con las vecinas, pues es sabido que la conductividad térmica del aire es muy baja (de ahí el uso de las "cámaras de aire" por los arquitectos con fines de aislamiento térmico en edificios) y en cualquier caso, el proceso de conducción molecular en gases sería lento. El único medio eficaz que tiene el aire para tomar o ceder calor es el intercambio turbulento en pequeña escala que se menciona más adelante. Este, sin embargo, requiere también cierto tiempo y afecta a la zona cortical de la partícula, que si es suficientemente grande, conservará siempre un núcleo no afectado por tal intercambio. En estas condiciones la partícula evoluciona esencialmente sin intercambiar calor ni materia con el aire circundante, clase de evolución que se llama adiabática; en ella sólo queda una fuente posible de donde la partícula puede tomar la energía requerida para su dilatación: de sí misma. En efecto, la partícula (como todo gas) mantiene una especie de reserva de energía llamada energía interna, que en primera aproximación es igual a la suma de las energías cinéticas de las moléculas del gas al efectuar su movimiento casi perfectamente desordenado. Se puede demostrar que la energía interna de un gas ideal (aproximación suficiente para la atmósfera) es proporcional a su temperatura absoluta (es decir, expresada en grados Kelvin). Consiguientemente, si la partícula en su ascenso necesita consumir parte de su energía interna para dilatarse, su temperatura disminuirá: es decir, la partícula se enfría al ascender adiabáticamente, pero no porque el aire circundante esté más frío sino porque ella ha tenido que realizar un trabajo de dilatación a costa de su propia energía interna.

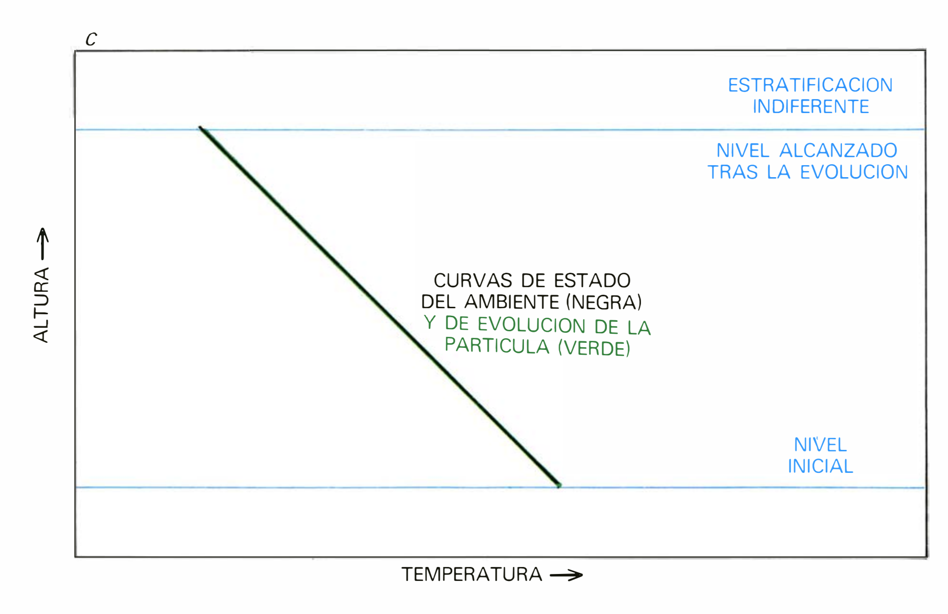
El decrecimiento de temperatura de la partícula al elevarse adiabáticamente una unidad de longitud puede calcularse fácilmente: cuando no hay condensación del vapor de agua resulta valer casi exactamente un grado Kelvin por cada 100 metros de ascenso. Se llama coeficiente de enfriamiento por elevación adiabática seca (con mucha frecuencia, aunque impropiamente, "gradiente adiabático seco de temperatura") y suele designarse por γ.

Resulta que tanto en el aire ambiente como en la partícula que evoluciona adiabáticamente sin mezclarse con él, observamos una disminución de temperatura al aumentar la altura. Ocurre sin embargo, y éste es el punto crucial de la cuestión, que en general la partícula no se enfría en la misma cuantía que el ambiente. Este lo hace según el valor variable de α; la partícula, mientras no haya condensación, según el valor constante γ. La diferencia es la clave del carácter de la estratificación.

Consideremos el caso habitual en la atmósfera en que es α menor que γ, esto es, al elevarse una unidad de longitud la partícula queda más fría que el ambiente. Si al iniciarse el ascenso ambos







tenían la misma temperatura (es decir, la partícula formaba parte del ambiente) se produce una situación como la representada en la ilustración (*a*) de esta página: al final de su ascenso la partícula, más fría que el aire circundante a la misma altura, es por tanto más densa que él. Consiguientemente, tenderá (si otras fuerzas no lo impiden) a regresar a su nivel de partida a fin de que el fluido más denso se encuentre abajo. Se dice entonces que la estratificación de la atmósfera es estable. Obsérvese que si en lugar de ascender, se hubiera forzado a la partícula a descender partiendo del punto A [*véase dicha ilustración*], el resultado hubiese sido el mismo: la partícula en su descenso adiabático hubiera llegado al nivel C más caliente, y, por tanto, menos densa que el aire ambiente y hubiera tendido a subir nuevamente hasta A. La estabilidad de estratificación implica la tendencia de cualquier partícula accidentalmente separada de su nivel de equilibrio a regresar a él.

El caso opuesto sería aquel en que el

PENDIENTE DE LA CURVA DE ESTADO. Ella nos determina el carácter de la estratificación atmosférica en la vecindad de un punto. Cuando una partícula de aire se separa de su nivel inicial, evoluciona esencialmente sin intercambiar calor ni materia con el aire circundante, evolución que se llama adiabática. Al elevarse, la partícula se dilata y efectúa trabajo a costa de su propia energía interna; consiguientemente se enfría. El enfriamiento por cada 100 metros de elevación vale casi exactamente un grado Celsius. Según cuál sea la distribución de temperatura en el ambiente (curva de estado), la partícula puede alcanzar su nuevo nivel a temperatura distinta de la del ambiente. En la figura de arriba, la curva de estado queda a la derecha de la curva de evolución adiabática, lo que significa que, en el nuevo nivel, la partícula está más fría que el aire circundante y por tanto es más densa que él. La partícula tenderá a regresar a su nivel de partida y por ello se dice que la estratificación es estable. Cuando la curva de estado queda a la izquierda de la curva de evolución (*centro*), la partícula alcanza su nuevo nivel más caliente y por ello menos densa que el ambiente y tiende a continuar ascendiendo, es decir, a alejarse más de su nivel de partida. La estratificación en este caso se llama inestable. Si la temperatura de la partícula y la del ambiente coinciden en todo el curso de la evolución (*abajo*), a cualquier nivel distinto del inicial la partícula está a la temperatura ambiente y tiene por tanto la misma densidad que el aire circundante. No tiende a seguir subiendo ni a bajar, y por ello se habla en este caso de estratificación indiferente. Obsérvese que de las tres clases de estratificación, la estable es la única que tiende a frenar los movimientos verticales, obligando a las partículas separadas de su nivel inicial a regresar a él. El efecto es tanto más marcado cuanto más estable la estratificación, es decir, cuanto más se separen las curvas de estado y evolución en la figura de arriba. Las inversiones constituyen un caso extremado de estabilidad y, por tanto, inhiben eficazmente los movimientos verticales de la atmósfera. Como la dispersión de contaminantes se realiza en gran parte gracias a estos movimientos, se deduce que la presencia de inversiones en la parte baja de la troposfera es desfavorable.

enfriamiento geométrico del ambiente, α, fuese mayor que el de la partícula en su ascenso adiabático, γ. Esta condición se da frecuentemente en latitudes medias o bajas en verano y junto al suelo, cuando éste es fuertemente calentado por el sol. El aire junto al suelo es a su vez calentado por éste y la curva de estado del ambiente adopta un aspecto parecido al de la ilustración *b* de la página 110. Debemos señalar, sin embargo, que esta situación sólo rarísima vez se da en la atmósfera libre, lejos de la influencia del suelo. Una partícula que inicialmente se encontrase junto al suelo podría en este caso iniciar espontáneamente su ascenso. Una vez perdido el contacto con el suelo, el ascenso de la partícula será, como antes, adiabático, pero merced a la peculiar distribución vertical de temperatura del ambiente, a una cierta altura la partícula (*curva en color*) va a estar más caliente que él, siendo por tanto menos densa y consiguientemente tenderá a seguir ascendiendo, alejándose de su nivel original. Una partícula tomada del ambiente con esta curva de estado (punto P) a la que se forzase a descender, se calentaría adiabáticamente y a cualquier nivel por debajo del inicial se encontraría más fría, y por ello más densa, que el ambiente y tendería a seguir descendiendo, nuevamente alejándose del nivel de partida. Se dice en este caso que la estratificación de la atmósfera es inestable.

Finalmente, puede ocurrir que un tramo de la curva de estado coincida con la curva de evolución adiabática que pasa por allí, es decir, que se tenga $\alpha = \gamma$. En la vecindad del suelo esto se da con cielo cubierto y viento moderado a fuerte. En estas circunstancias, cualquier partícula separada de su nivel se enfría en su ascenso y se calienta en su descenso en la misma proporción en que varía la temperatura del ambiente. En otras palabras, a cualquier nuevo nivel la partícula se encuentra a la misma temperatura del ambiente; no será más ni menos densa que él, y carecerá de toda tendencia a seguir subiendo o a bajar. Este es el caso de estratificación indiferente [*caso c de la figura citada*].

Se deduce de lo anterior que una atmósfera con estratificación estable tiende a suprimir las corrientes verticales porque las partículas que se separan de su nivel son rechazadas de nuevo hacia él, y ello sucede de modo tanto más eficaz cuanto mayor es la estabilidad. Las inversiones, que constituyen casos extremos de estratificación estable, actúan a modo de barrera infranqueable para las corrientes verticales. Por el contrario, éstas son extraordinariamente activas en ambos sentidos en el caso de estratificación inestable. Cuando la estratificación es indiferente, no existe tendencia por parte de la atmósfera a favorecer tales corrientes, pero tampoco a suprimirlas si se producen (pueden producirlas la orografía o la turbulencia); por tanto las corrientes verticales se dan frecuentemente con este carácter de estratificación. Cuanto mayor sea la frecuencia e intensidad de las corrientes verticales, tanto más fácil será la dispersión o reparto vertical de contaminantes; en ausencia de viento apreciable, este es de hecho el único mecanismo de que la atmósfera dispone para reducir su concentración. La estratificación estable, que tiende a suprimir tales corrientes, es la más desfavorable a este respecto, y tanto más cuanto más estable. La presencia de inversiones junto al suelo o a baja altura suele ser indeseable.

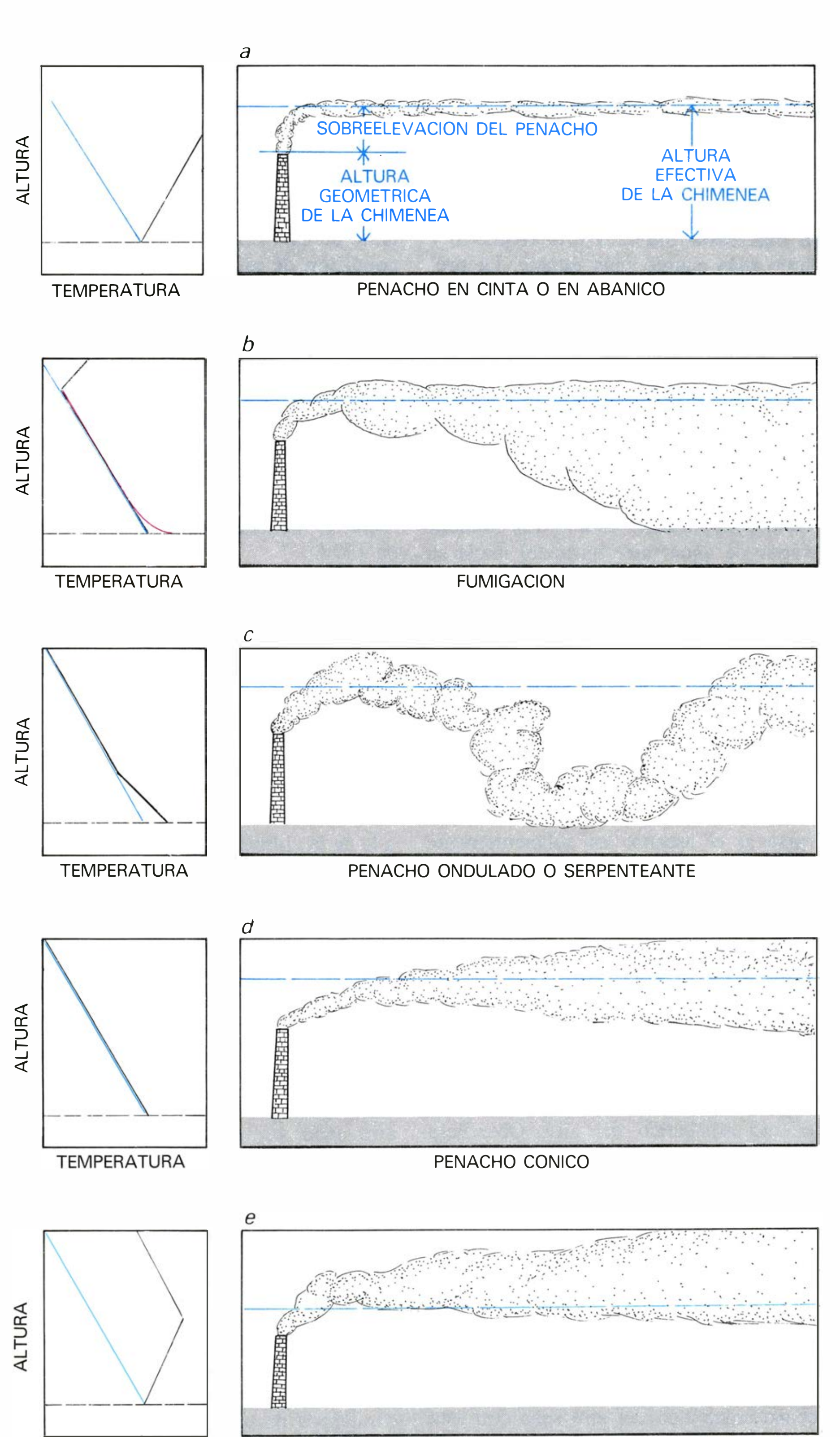
La interpretación anterior es ciertamente simplista porque hemos prescindido del efecto del vapor de agua sobre la partícula ascendente. El aire siempre contiene algo de vapor de agua; el ascenso adiabático seco es una idealización. Una partícula de aire que contiene vapor de agua (aire húmedo) inicia su ascenso de manera similar al caso del aire seco, pero a medida que sube y se enfría, su humedad relativa va aumentando. (La humedad relativa es el cociente entre la masa de vapor de agua que contiene la partícula y la que podría contener si estuviese saturada a la misma temperatura.) Puede demostrarse que cualquiera que fuese la humedad relativa inicial, la partícula llega a saturarse si el ascenso es suficientemente prolongado. A partir de la saturación, el vapor comienza a condensarse si el ascenso –y el enfriamiento– prosigue.

(De nuevo el proceso es más complicado, pero lo dicho basta para nuestro propósito.) Sin embargo, en la condensación el vapor pone en juego una importante cantidad de calor (casi 600 calorías por gramo de vapor condensado) que se invierte en atenuar el enfriamiento adiabático del aire. Obsérvese que en este caso el ascenso ya no es adiabático para el aire de la partícula, que recibe el calor de condensación del vapor, pero la masa del primero es enormemente superior a la de vapor (la relación entre ambas, llamada proporción de mezcla, se suele expresar en gramos de vapor por kilo de aire y su valor oscila entre 40 y 0,001) y por ello no hay grave inconveniente en comparar el ascenso adiabático del aire seco con éste en que el vapor se condensa, llamado ascenso adiabático saturado. El coeficiente de enfriamiento por unidad de elevación adiabática saturada (vulgarmente llamado, no del todo correctamente, gradiente adiabático saturado Γ) ya no es constante ni vale un grado C por 100 metros: es siempre menor, y su valor depende de la cantidad de vapor condensado, que a su vez depende de la temperatura y la proporción de mezcla inicial. En regiones tropicales al nivel del suelo, llega a valer 0,4 grados C por 100 m, mientras que a 10 km y −60 grados C vale casi lo que γ: a medida que la temperatura va disminuyendo, también lo hace la cantidad de vapor que puede condensarse, y el gradiente adiabático saturado Γ se va aproximando al seco, γ. Consecuencia importante de este comportamiento es que la condición de estabilidad para aire saturado $\alpha < \Gamma$ es más difícil que la $\alpha < \gamma$ del aire seco, pudiendo darse la condición $\Gamma < \alpha < \gamma$: la estratificación será estable mientras el aire no se sature e inestable cuando lo haga. Estos resultados tienen evidente interés tanto en relación con el carácter de la estratificación atmosférica como el análisis del comportamiento de penachos húmedos descrito en el número de *Investigación y Ciencia* de julio de 1978.

El efecto de la variable distribución de α a lo largo de la curva de estado sobre la forma del penacho de humo de una chimenea se esquematiza en la ilustración de la página 112, en cuya parte izquierda se muestra la porción relevante de la correspondiente curva de estado. En el caso (*a*) existe una inversión desde el suelo hasta una altura superior a la de la chimenea. Esta situación se da como consecuencia del enfriamiento nocturno del suelo y de la capa de aire junto a él por emisión de radiación de gran longitud de onda, la llamada radiación térmica, infrarroja o terrestre. Si la duración de la noche es suficiente y el cielo está despejado (pues las nubes actúan a manera de manta, entorpeciendo efectivamente el enfriamiento nocturno), el enfriamiento se extiende a capas de aire cada vez más espesas al mismo tiempo que pierde intensidad cuanto más arriba. Para el desarrollo de la inversión conviene que el viento sea muy flojo o en calma; de otra manera, la turbulencia reparte el enfriamiento en capas más espesas y en vez de inversión simplemente se reduce el gradiente vertical de temperatura. Bajo las condiciones señaladas (lo que sucede preferentemente a fines de otoño, en invierno y a comienzos de primavera), es decir, con la curva de estado del tipo (*a*), el movi-

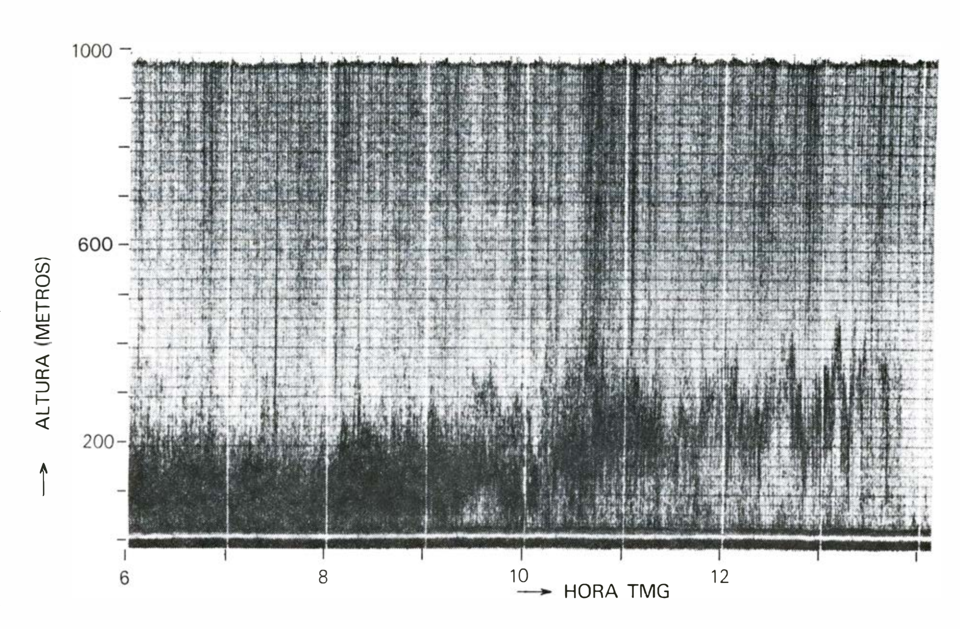
miento vertical está prácticamente eliminado; el penacho se extiende horizontalmente en forma de cinta, si no hay dispersión lateral, o de abanico cuando la hay, mientras que verticalmente apenas se difunde.

Las inversiones de superficie se destruyen generalmente mediante un mecanismo contrario al de su formación. Desde la salida del sol comienza el calentamiento del suelo por la radiación solar, que sin embargo no calienta el aire libre (recordemos que éste es esencialmente transparente respecto a la radiación solar). En cambio, la delgada capa de aire inmediata al suelo se va calentando por absorción de radiación de onda larga emitida por el suelo. Las partículas de esta capa acaban por adquirir menor densidad que las que tienen inmediatamente encima y tienden a ascender mientras que las otras descienden. Como el proceso es lento y las partículas que en él toman parte son de pequeñas dimensiones, el intercambio de calor y materia entre unas y otras es muy activo, con el resultado de que en la delgada capa afectada tiende a establecerse gradiente vertical de temperatura igual al adiabático [*véase la ilustración (b) de esta página*]. Si las condiciones son favorables para que el proceso continúe (es decir, cielo despejado o casi despejado y viento muy flojo o en calma), las capas de aire afectadas por él van siendo cada vez más espesas hasta la total destrucción de la inversión, que queda reemplazada por una capa de gradiente próximo al adiabático seco. En el caso (*b*) de la figura, la inversión ha sido destruida



LA FORMA QUE ADQUIERE EL PENACHO de una chimenea puede servir para deducir el carácter de la estratificación en la parte más baja de la atmósfera. La correspondiente curva de estado (*en negro*) se esquematiza en los diagramas de la izquierda, donde también se representa la curva de evolución adiabática (*color*). La figura de arriba (*a*) muestra el caso de una inversión que parte del suelo y alcanza una altura bastante superior a la de la chimenea. El humo que sale de ésta asciende al principio (casi verticalmente, pues en esta situación hay poco o ningún viento) porque los humos salen en general calientes y con una cierta velocidad debida al tiro forzado. Alcanzada cierta altura (sobreelevación del penacho), la fuerza ascensional se ha anulado. Como la inversión impide las corrientes verticales, el penacho se extiende en forma de cinta si no hay dispersión lateral o de abanico si la hay; en ambos casos la dispersión vertical es muy pequeña. En la segunda figura desde arriba (*b*), se supone que la inversión de superficie ha sido destruida por calentamiento hasta una altura un poco superior a la altura efectiva de la chimenea (*representada en todas las figuras por la línea azul de trazos*). La base de la inversión impide la dispersión del penacho hacia arriba, pero la estratificación indiferente por debajo (o incluso inestable si el calentamiento es suficiente, *curva roja*) permite o estimula las corrientes verticales, con lo que el penacho se difunde hacia abajo llegando a tocar el suelo. Este caso, llamado fumigación, es el que produce mayores concentraciones de contaminantes en el suelo. Si el calentamiento es insuficiente para destruir por completo la inversión, las concentraciones pueden alcanzar valores peligrosos. En la tercera figura, (*c*), la inversión ha sido completamente destruida y las corrientes verticales, muy activas, obligan al penacho a adoptar formas cambiantes. El penacho puede ocasionalmente alcanzar el suelo y producir en algún lugar contaminación alta, pero se trata de un episodio de poca duración. La situación representada en la cuarta figura, (*d*), corresponde a estratificación indiferente. En la vecindad del suelo, este caso se da acompañado de cielo muy nuboso o cubierto y vientos moderados o fuertes y racheados. El penacho se extiende hacia arriba y hacia abajo, pero el efecto del viento predomina y le hace adoptar una forma esencialmente cónica. El caso de la última figura, (*e*), es en cierto modo inverso al (*b*). Se produce al comienzo de la noche, cuando el enfriamiento nocturno da lugar a una inversión de superficie todavía no muy espesa. En la figura se ha supuesto su espesor algo inferior a la altura efectiva de la chimenea. La ausencia de movimiento vertical en la región de inversión impide al penacho extenderse hacia abajo, en tanto que nada inhibe su crecimiento hacia arriba.

hasta una altura un poco mayor que la de la chimenea y ha sido reemplazada por una región de pendiente próxima a la de la adiabática seca. En estas condiciones existirá fuerte agitación vertical por debajo de la base de la inversión y poca o ninguna por encima de ella. El penacho y los contaminantes quedan aprisionados entre la base de la inversión y el suelo, y a cierta distancia de la chimenea pueden alcanzarlo, produciéndose el fenómeno llamado de fumigación: esta es la situación más desfavorable en cuanto a concentración de contaminantes en el suelo. En la ilustración de la derecha se presenta un caso de fumigación puesto de manifiesto mediante un método acústico de exploración remota de la baja atmósfera conocido por SODAR: la banda negra de la gráfica entre el suelo y unos 200 metros corresponde a la inversión nocturna, que comienza a ser erosionada por debajo a partir de las 9,30 (hora de Greenwich) hasta desaparecer hacia las 14; entre ambas horas existen condiciones propicias a la fumigación. Pese a ser desfavorable, una situación como la descrita no es crítica a causa de su duración limitada; puede serlo, sin embargo, si el calentamiento no llega a destruir la inversión en el curso del día o si se trata de otra clase de inversión más persistente, llamada de subsidencia, que se describe abajo.



SITUACION PROPICIA a la fumigación, detectada mediante un equipo de exploración remota de la atmósfera conocido por SODAR (*So*und *De*tection *a*nd *R*anging). El método consiste en enviar a la atmósfera cortos trenes de ondas sonoras de frecuencia fija y fuertemente colimados y detectar en tierra mediante un equipo adecuado la porción de energía sonora retrodifundida por la atmósfera. La teoría muestra y la experiencia comprueba que en el caso de haz dirigido verticalmente (Sodar monostático) la retrodifusión es debida a irregularidades en la distribución de la temperatura exclusivamente. El registro suele obtenerse en la forma de diagrama altura-tiempo. En el caso aquí presentado, obtenido por el autor y colaboradores en Barcelona, la región negra y ligeramente ondulada junto al suelo hasta unos 200 metros corresponde a una inversión nocturna que se va elevando y debilitando a partir de las 9,30 horas del meridiano de Greenwich hasta quedar completamente destruida hacia 14 horas. El período comprendido entre el momento en que la inversión comienza a perder contacto con el suelo y aquel en que queda destruida es favorable para la fumigación aunque ésta no se produzca necesariamente. La destrucción de la inversión nocturna fue efectuada por el calentamiento solar del suelo y el aire junto a él; la situación en ese período corresponde aproximadamente a la representada en el caso (*b*) de la ilustración dibujada en la página anterior (inversión solamente por encima de la chimenea).

Volviendo a la ilustración de la página 112, en el caso (*c*) la inversión ha sido completamente destruida y el fuerte calentamiento del suelo por los rayos solares crea junto a éste una región de gradiente superadiabático, siendo el resto aproximadamente igual al adiabático. Las corrientes verticales son extremadamente activas, tanto ascendentes como descendentes, y el penacho se ondula adquiriendo un aspecto serpentino. Eventualmente puede alcanzar el suelo, produciendo un nivel alto de contaminación, pero se trata de una situación transitoria porque el penacho cambia de forma continuamente y en general la velocidad de dispersión es elevada.

El caso (*d*) corresponde a un gradiente vertical de temperatura próximo al adiabático seco y suele darse en días nubosos o cubiertos y con viento moderado a fuerte; el penacho se difunde al mismo tiempo que es arrastrado por el viento y adquiere forma cónica.

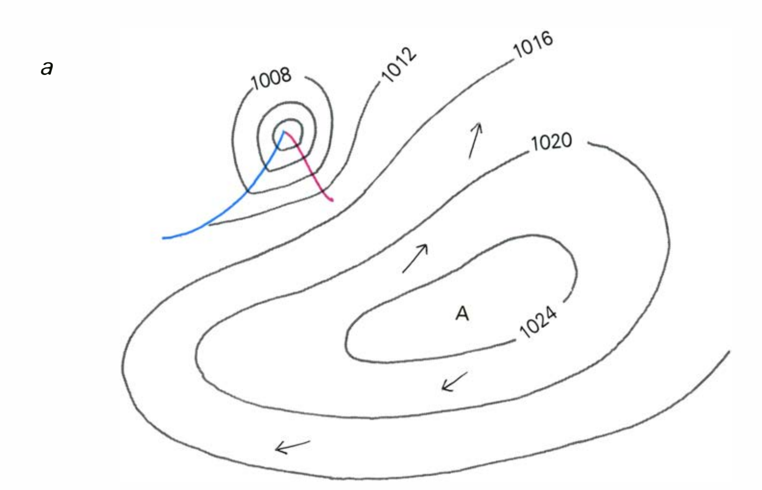
El caso (*e*) es en cierto modo inverso al (*b*): se está formando una inversión de superficie de espesor todavía pequeño, aproximadamente igual a la altura de la chimenea. Ello se produce tras la puesta del sol en un día despejado, al comienzo del enfriamiento nocturno. La inversión impide la difusión de contaminantes hacia abajo pero no hacia arriba y se tiene un penacho ascendente. Si el enfriamiento es suficiente para que el espesor de la inversión supere a la altura de la chimenea, volvemos a estar en (*a*).

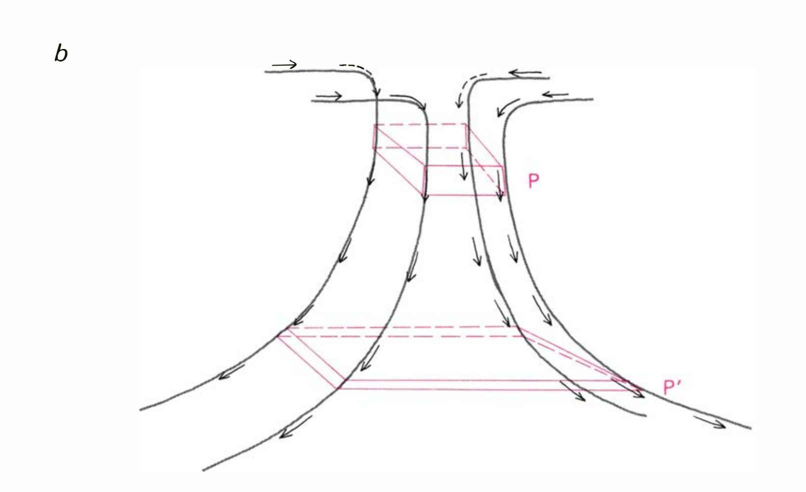
El otro mecanismo de formación de inversiones que es importante a nuestros efectos, pues puede producir fumigación persistente, es de carácter dinámico y se conoce por subsidencia. Es característico de los anticiclones o sistemas de alta presión, que se extienden, cuando están bien desarrollados, sobre áreas de millones de kilómetros cuadrados. En ellos, el viento a niveles bajos tiende ligeramente a escapar del centro (divergencia), movimiento que está compensado por una circulación en el plano vertical consistente en un movimiento lento y persistente de descenso en bloque del aire (subsidencia), típicamente unos 1000 m por día, particularmente apreciable en la zona central del anticiclón y entre aproximadamente 5000 metros y 500 metros [*véase la ilustración de la página 114, a* y *b*]. El efecto de la subsidencia sobre la curva de estado del ambiente puede apreciarse en la misma ilustración, (*c*): el descenso produce compresión y por tanto calentamiento adiabático, de modo que la partícula P pasa en un par de días al estado P'. La proximidad del suelo entorpece la subsidencia, por lo que el aire allí desciende mucho menos; también lo hace a niveles muy altos, donde la convergencia inicia su transformación en subsidencia [*véase la misma ilustración, b*]. El resultado es una curva de estado (*color azul*) que presenta una fuerte inversión con base a unos 500 metros mientras que por encima de ella la estratificación es menos estable que en la curva gris inicial. La curva de color negro, que representa la fase final del proceso, se alcanza por enfriamiento radiativo del aire en el estado representado por la curva azul; el aire de la inversión, más caliente que el de arriba y el de abajo emite radiación infrarroja y se enfría apreciablemente. En esta fase final, el calentamiento por subsidencia y el enfriamiento radiativo se compensan mutuamente y la curva de estado evoluciona poco.

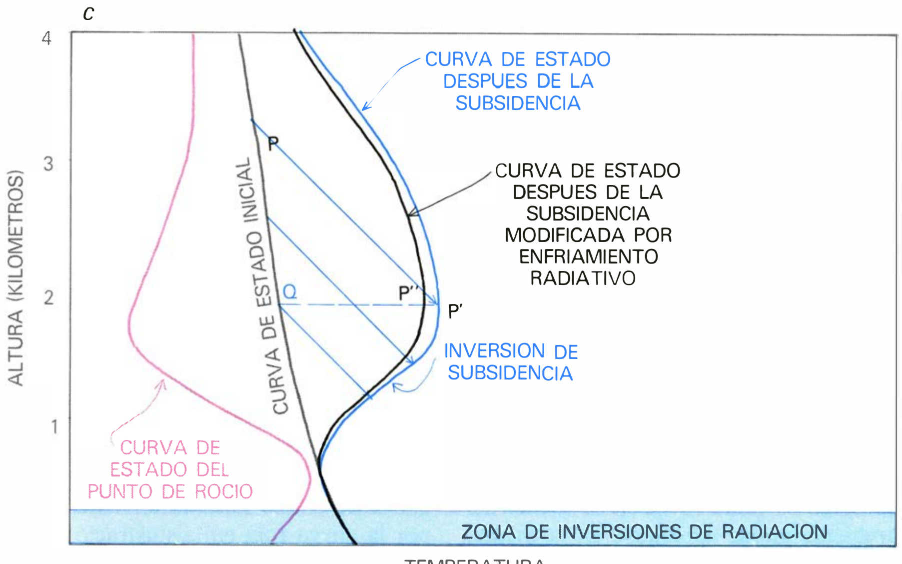
El mismo calentamiento adiabático que origina la inversión de subsidencia produce en ella un fuerte y brusco descenso de la humedad relativa, cuya curva de estado toma la forma indicada por la curva roja en la ilustración de la página 114, (*c*). Como la subsidencia

suele dar lugar a cielos despejados, el calentamiento del suelo puede fácilmente originar convección que es detenida por la inversión, con el resultado de que en la base de ésta se acumula el vapor de agua, los núcleos de condensación y las partículas contaminantes. La base de la inversión puede fácilmente saturarse, formándose en ella una capa nubosa. Esta emite radiación infrarroja de forma mucho más eficaz que el aire; con ello su cima se enfría mientras que la base se comporta de manera complicada, actuando esencialmente como una manta que impide o mitiga el enfriamiento del suelo. Consiguientemente, la inversión se hace más abrupta al mismo tiempo que el gradiente geométrico del aire bajo la inversión se aproxima al adiabático seco, dando por resultado una región de mezcla muy activa coronada por una fuerte inversión.

Cuando se inyectan contaminantes en el aire bajo la inversión, estos son repartidos en toda la capa por los movimientos verticales de ascendencia, pero no pueden pasar más arriba de la base de la inversión. Cuando el anticiclón se presenta en invierno, a los efectos de la subsidencia se añaden los de enfriamiento nocturno que pueden originar inversiones de superficie. Añádase la escasa ventilación (los vientos cerca del suelo son siempre flojos en los anticiclones), el hecho de que los anticiclones sean sistemas de vida relativamente larga (a veces va-



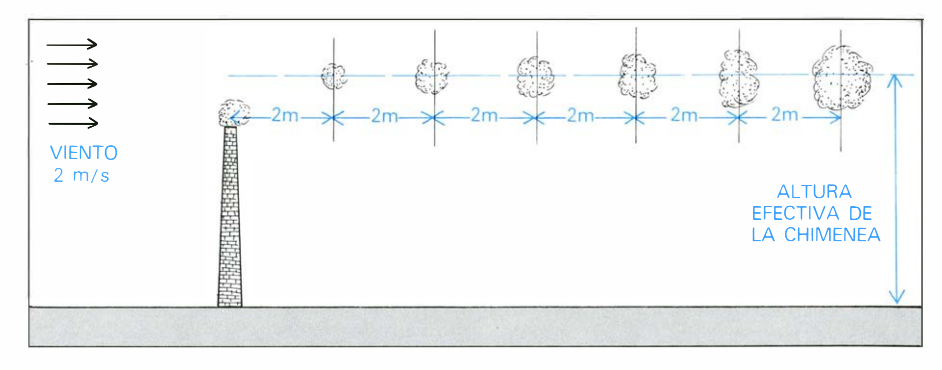


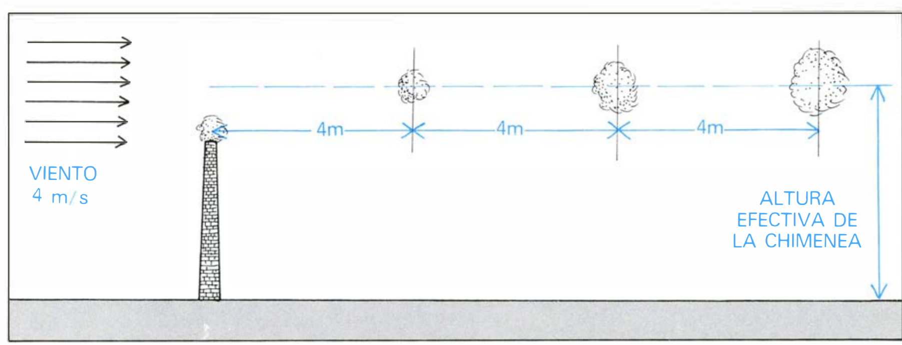


UN ANTICICLON (*arriba*) es una zona de alta presión atmosférica que se extiende horizontalmente sobre miles de kilómetros cuadrados y verticalmente con frecuencia hasta la base de la estratosfera. En el Hemisferio Norte, los vientos en el anticiclón circulan en torno a la parte central en el sentido de las agujas del reloj, pero no siguen estrictamente las isobaras (*curvas negras*, rotuladas en hectopascal) sino que forman un pequeño ángulo con ellas, cruzándolas hacia la periferia. Los vientos, que en la parte central del anticiclón son siempre flojos, están representados mediante flechas en la figura. La divergencia originada por los vientos hacia fuera en superficie se compensa en la parte alta del anticiclón (generalmente algo por debajo de la tropopausa) mediante convergencia y un lento movimiento de descenso continuo en la parte central del anticiclón, llamado subsidencia. A causa de este movimiento (*figura intermedia*) un estrato de aire limitado por un prisma (*en color*) va descendiendo lentamente al mismo tiempo que se aplasta en sentido vertical (por ir encontrando presiones cada vez mayores) y se dilata en sentido horizontal por efecto de la forma de las líneas de corriente. La compresión adiabática produce calentamiento del estrato. El descenso se hace cada vez menos perceptible cuanto más cerca del suelo. El efecto de la subsidencia sobre la curva de estado se muestra en la figura (*c*). Una partícula situada inicialmente en el punto *P* desciende al mismo tiempo que se calienta por efecto de la subsidencia, alcanzando (típicamente en día y medio o dos días) el estado representado por el punto *P'*, mucho más caliente que el punto *Q*, que en el sondeo inicial correspondía a la misma altura. Ulteriormente, el enfriamiento por radiación puede conducir a la partícula al estado *P''*, todavía mucho más caliente que el *Q*. Haciendo la construcción punto a punto se obtiene la curva de estado azul: la subsidencia de los estratos que inicialmente estaban más bajos es menor por el entorpecimiento que el suelo ejerce sobre el descenso, y consiguientemente también es menor el calentamiento que experimentan. El resultado es la formación de una fuerte inversión a alguna distancia del suelo (600 a 800 metros). Además, el calentamiento adiabático producido por el descenso del aire reduce su humedad relativa; la variación con la altura de ésta o de otro índice de humedad tiene el aspecto de la curva roja: la inversión de subsidencia se caracteriza por un fuerte descenso de la humedad al pasar de la base a la cima de la inversión. Estas inversiones, por su persistencia y porque van descendiendo a lo largo de la vida del anticiclón, constituyen un ingrediente casi indispensable en los episodios catastróficos de contaminación.

rias semanas) y la tendencia de la inversión de subsidencia a ir descendiendo a lo largo de la vida del anticiclón (puede formarse a 600 u 800 m y acabar su vida a unos 200 m sobre el suelo). Todo ello configura un conjunto de condiciones propicias para elevados valores de la contaminación que pueden persistir lo suficiente para que la acumulación de contaminantes llegue a crear situaciones de alarma o de franco peligro. De hecho, algunos de los más catastróficos episodios de contaminación conocidos se produjeron en situaciones análogas a las que acaban de describirse (Valle del Mosa, Bélgica, 1 a 5 de diciembre de 1930; Donora, Pennsylvania, 26 a 29 de octubre de 1948; Londres, 5 a 9 de diciembre de 1952; New York, 24 a 30 de noviembre de 1966).

Un último punto merece mención antes de dar por terminado este apartado. Se ha razonado más arriba que la estratificación estable es desfavorable a efectos de contaminación, lo que puede confundir a algunos lectores que, al menos en España, se han acostumbrado a que les pronostiquen "estabilidad" como sinónimo de buen tiempo. Tratando de la atmósfera, la estabilidad puede entenderse en el sentido hidrostático de carácter de la estratificación arriba descrito, o en sentido hidrodinámico, hasta ahora no mencionado porque carece de relación con nuestro problema. En este caso, la inestabilidad se entiende como la condición que favorece la amplificación de una pequeña perturbación superpuesta a una corriente en gran escala. En ninguno de los dos casos se prejuzga una relación unívoca con el tiempo, cosa mucho más complicada, ni aunque existiera sería normal que el predictor la empleara en un informe para el público. El inexplicable y persistente uso de locuciones sin sentido como "estabilidad atmosférica" para designar buen tiempo e "inestabilidad" para indicar el malo o el variable (pues ni siquiera esto queda claro) por parte de algunos predictores constituye un buen ejemplo de práctica viciosa que debería eliminarse.



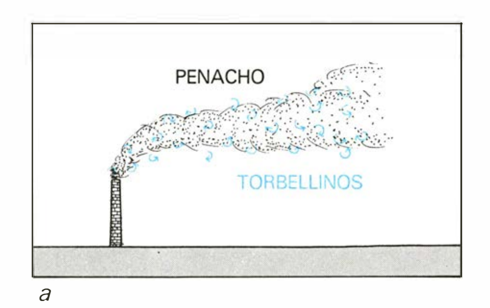


EFECTO DE LA VELOCIDAD del viento sobre la concentración de contaminantes en el penacho. Una chimenea hipotética emite una bocanada de humos cada segundo. Si el viento es de dos metros por segundo (*arriba*), en el momento de ser emitida una bocanada, la anterior se encontrará a dos metros de la boca de la chimenea, la que precede a ésta a cuatro metros, y así sucesivamente. Si el viento fuera de cuatro metros por segundo (*abajo*), la distancia entre las sucesivas burbujas en las mismas condiciones sería de cuatro metros y la cantidad total de contaminantes a lo largo de una distancia dada según la altura efectiva se habría reducido a la mitad con respecto al caso de arriba. La concentracíon medida resulta ser, por tanto, inversamente proporcional al módulo de la velocidad que ostenta el viento.

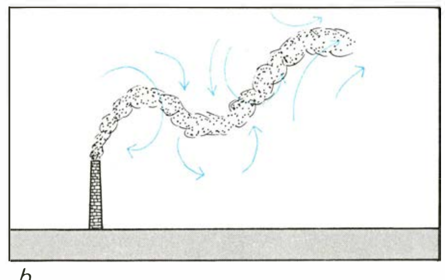
El segundo factor meteorológico de importancia fundamental en lo que concierne a la difusión (o dispersión) de contaminantes es el viento. Para comprender bien el efecto del viento hay que guardarse de considerarlo como una corriente que fluye suavemente a lo largo de invisibles canales atmosféricos. Esta analogía puede parecer pueril; pero cuando se informa que el viento sopla de 220 grados y 16 nudos, pocos piensan que hay "algo más". (Desde los tiempos de la navegación a vela, la dirección del viento se da por la orientación en la brújula del punto de donde procede. El nudo es una milla náutica por hora y la milla náutica es la longitud del arco de círculo máximo terrestre que subtiende un ángulo de un minuto; vale 1852 metros. Actualmente se usa ya el metro por segundo –aproximadamente 0,5 nudos– en trabajos científicos, pero pasará tiempo antes de que los marinos y aviadores abandonen el nudo, más conveniente a sus efectos). Sin embargo, quien haya observado el flameo de una bandera convendrá en que si bien en términos generales el viento viene de un punto del horizonte, presenta fluctuaciones más o menos amplias en torno a esa dirección básica. Lo propio sucede con la velocidad, y cuando el viento es moderado o fuerte puede observarse simplemente mirando las hojas de los árboles: existen rachas y calmas parciales superpuestas a la velocidad media del viento. (Cuando éste es flojo, la detección de tales fluctuaciones puede requerir el uso de un instrumento sensible.) En resumen, al viento "medio" se superponen fluctuaciones instantáneas de amplitud variable tanto en dirección como en velocidad. En los instrumentos usados en las observaciones meteorológicas corrientes se procura eliminarlas dentro de ciertos límites dotándolos de alguna inercia para obtener los datos de la corriente básica; el usuario profesional (es decir, marinos, aviadores y meteorólogos) ya espera la variabilidad superpuesta a esos valores.

Sin embargo, en problemas relacionados con la estructura fina de la atmósfera (lo que algunas veces se llama micrometeorología, que incluye la meteorología de la contaminación atmosférica), interesan fundamentalmente esas fluctuaciones, que son en realidad movimientos aleatorios o caóticos en las tres direcciones y en distintas escalas de tamaño y tiempo y forman parte integrante del movimiento constituyendo la turbulencia, probablemente la más difícil rama de la Mecánica de fluidos.

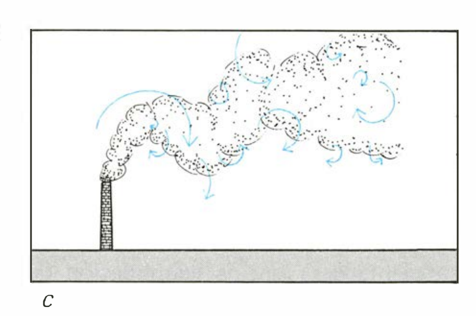
Para estimar la influencia del viento sobre la concentración de contaminantes, volvamos por un momento al tosco modelo de viento de dirección y velocidad constante e imaginemos una chimenea que emite bocanadas discretas de humo, una cada segundo. La ilustración de esta página permite darse cuenta de cómo la duplicación de la velocidad del viento reduce a la mitad la concentración de contaminantes. Por tanto, a igualdad de otras condiciones debe esperarse que también en el caso de emisión continua, la concentración de contami-



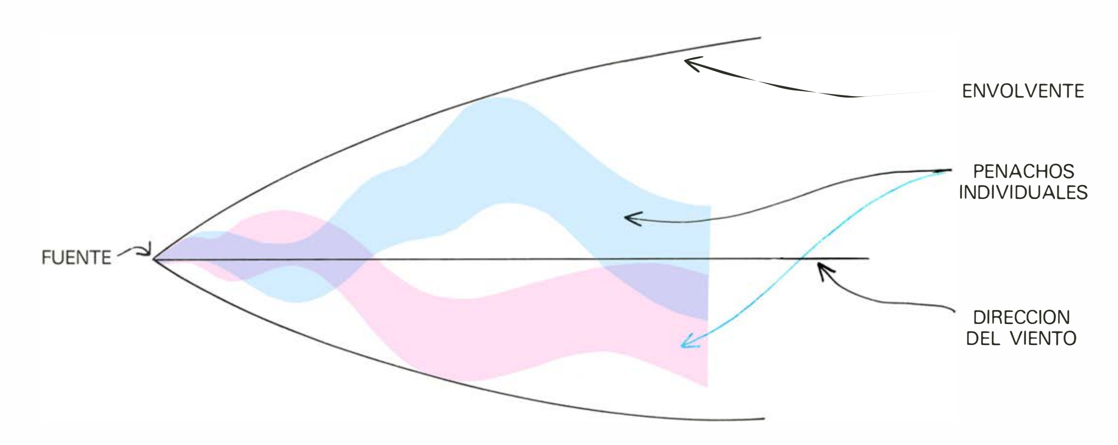
a



b



c



LOS TORBELLINOS de pequeñas dimensiones (*a*) dan por resultado aumentar ligera y gradualmente la sección del penacho; los torbellinos de dimensiones mucho mayores que las del penacho (*b*) hacen serpentear éste sin aumentar apreciablemente su sección. Como en la atmósfera existen frecuentemente torbellinos de muy diversas escalas, los hay también de dimensiones aproximadamente iguales a las de la sección del penacho y éstos son los más eficaces a efectos de difusión. El resultado (*c*) de la actuación simultánea de remolinos de diversas escalas es que el penacho se ondula al mismo tiempo que su sección aumenta. El penacho observado en una fotografía con exposición, o incluso a simple vista cuando las fluctuaciones son rápidas, es la envolvente de los diversos penachos instantáneos (*d*) representados en la figura de arriba por dos de ellos (en azul y en rojo).

nantes en un penacho decrezca corriente abajo en razón inversa de la velocidad media del viento. Pero la turbulencia no está nunca ausente en la atmósfera. Los torbellinos pueden ser producidos por convección térmica, por la irregularidad del suelo (incluyendo edificios) o por la rafagosidad del viento; su efecto es facilitar la mezcla del aire contaminado del penacho con el limpio de alrededor, tanto vertical como lateralmente, favoreciendo así la dispersión. La ilustración de arriba, *d*), explica cómo, en realidad, el penacho que observamos es la envolvente de los sucesivos penachos instantáneos deformados por remolinos turbulentos de tamaño semejante a la sección del penacho. Bajo ciertas hipótesis razonables puede mostrarse que cuando los torbellinos dominantes son mucho menores que las dimensiones del penacho [*a, en la ilustración*], la forma de éste debe ser paraboloidal; pero cuando los hay de diversos tamaños (*c*), la envolvente de los penachos es esencialmente cónica. En una sección perpendicular al eje del cono [*véase la ilustración de abajo*], la concentración de contaminantes es máxima en dicho eje y decrece simétricamente a uno y otro lado de él.

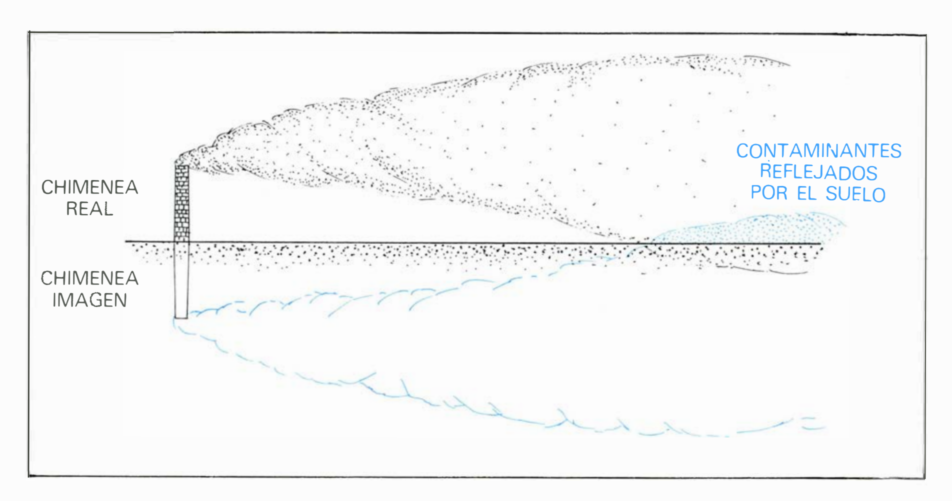
EN ESTE DIBUJO idealizado del penacho medido "envolvente" se ha representado la distribución de concentración de contaminantes a lo largo de dos ejes perpendiculares entre sí y al del penacho y situados en la misma sección recta de éste. Las medidas indican que la concentración es máxima en el eje del penacho y decrece hacia los bordes, pero no son suficientemente precisas para decidir qué función es la que mejor las representa. Una de las opciones posibles, que matemáticamente presenta ciertas ventajas, es la curva de campana de Gauss o curva de distribución de errores. Tal penacho se llama, por lo mismo, penacho gaussiano y es la base de la mayor parte de las fórmulas de difusión.

Se han propuesto muchas fórmulas para representar la distribución de la concentración de contaminantes en la sección recta del penacho, pero las medidas son demasiado imprecisas para suministrar una comprobación experimental.

Cualquier modelo que dé cuenta razonablemente bien de la anchura del penacho, la forma del perfil de concentración y la cantidad total de contaminantes a través de cada sección del penacho (igual a la emitida por la fuente en la unidad de tiempo, lo que se llama su "intensidad"), es aceptable. Quizá el modelo semiempírico más en boga es el llamado "penacho gaussiano", en el que se supone que la concentración de contaminantes según una sección transversal del penacho obedece a una ley análoga a la de distribución de errores o "curva de campana" de Gauss, en tanto que se acepta la dependencia inversa con la velocidad del viento para la concentración corriente abajo. Los parámetros que determinan la sección del penacho son difíciles de medir, pero Frank Pasquill, de la Sección de Baja Atmósfera del Meteorological Office británico, y D. Bruce Turner, de la Sección de Meteorología de la EPA, han dado métodos sencillos para estimarlos aproximadamente a par-

tir de observaciones meteorológicas corrientes, incluso cualitativas.

En la fórmula que se obtiene para el penacho gaussiano resulta también que la concentración en el suelo a una distancia dada de la fuente es inversamente proporcional al cuadrado de la altura de ésta sobre el suelo. De aquí la extraordinaria importancia que la altura de las chimeneas tiene en relación con la dispersión de contaminantes, hecho que desde hace mucho tiempo se toma en consideración en instalaciones industriales, pero parece estar olvidándose por lo que respecta a arquitectura urbana.



CUANDO SE MIDEN CONCENTRACIONES de contaminantes en el suelo a distancia suficiente de la fuente, se encuentran valores más altos que los predichos por la fórmula de difusión, en realidad próximos al doble de éstos. Ello se debe a que cuando el penacho alcanza el suelo (no necesariamente la porción visible del penacho), algunos contaminantes son absorbidos por éste, pero la mayor parte se reflejan y vuelven a la atmósfera. Para dar cuenta de este hecho se introduce en la fórmula del penacho gaussiano un retoque consistente en incluir una chimenea hipotética, imagen especular de la real con respecto al suelo, de idéntica intensidad que ésta (o masa de contaminantes emitida por unidad de tiempo). La concentración de contaminantes obtenida por cálculo en la zona sombreada es entonces la suma de la calculada con la fórmula inicial más la debida a la chimenea imagen, que es igual a la anterior. Este artificio permite obtener valores razonablemente concordantes con los observados.

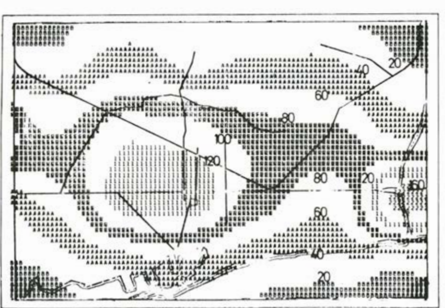
Por otra parte, los gases se expulsan, en general, con cierta velocidad y con frecuencia a temperatura más alta que la del ambiente [*véase Investigación y Ciencia, julio, 1978*]. Una y otra circunstancia dan por resultado que el eje del penacho llegue a quedar horizontal a una altura superior a la de la chimenea, que se llama su "altura efectiva" [*véanse las ilustraciones de las páginas 112 y 115*]. La diferencia entre dicha altura y la altura real es la llamada "sobreelevación del penacho", y en general no es despreciable.

Su conocimiento es de sumo interés en los cálculos de concentración de contaminantes mediante el modelo gaussiano, puesto que en la fórmula aparece el cuadrado de la altura efectiva. Para calcular la sobreelevación hace falta conocer la velocidad de salida de los gases, su temperatura, el diámetro de la chimenea y algún parámetro relacionado con el carácter de la estratificación atmosférica. Según la manera de considerar la influencia de las diversas variables resulta una abrumadora variedad de fórmulas, unas teóricas y otras empíricas, que pueden fácilmente confundir al presunto usuario. Se debe a Gary A. Briggs, del Laboratorio de Difusión y Turbulencia Atmosférica de Oak Ridge, dependiente de la Administración Nacional Oceánica y Atmosférica de los Estados Unidos (NOAA), una selección crítica que presenta las más adecuadas bajo distintas condiciones meteorológicas.

Muchos contaminantes emitidos se "reflejan" o "rebotan" en cuanto llegan al suelo, con el resultado de que el aire junto a éste queda más densamente contaminado de lo que indica la fórmula. La dificultad se remedia introduciendo una hipotética "chimenea imagen" [*véase la ilustración de arriba*] que en la práctica equivale a introducir un término adicional en la fórmula del penacho gaussiano que dé cuenta de la reflexión de contaminantes.

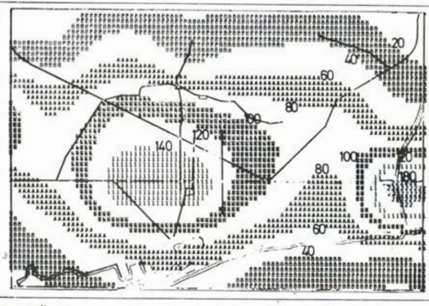
Pero lo anterior no puede aplicarse en ausencia de viento o en presencia de inversiones: como se ha visto, ambas condiciones afectan fundamentalmente la dispersión de contaminantes. El confinamiento del penacho bajo tales condiciones hace que el área de su sección recta crezca proporcionalmente a d (siendo d una dimensión transversal del penacho) y no a $d2$ como lo haría en ausencia de ellas. En consecuencia, el penacho se extiende lateralmente y su forma puede aproximarse mediante dos paraboloides adecuadamente escogidos. Turner ha indicado la manera de tratar aproximadamente estos casos introduciendo modificaciones en la fórmula del penacho gaussiano.

Con todo, subsiste la dificultad de que esta fórmula se refiere a una fuente puntual (es decir, de dimensiones desprecia-

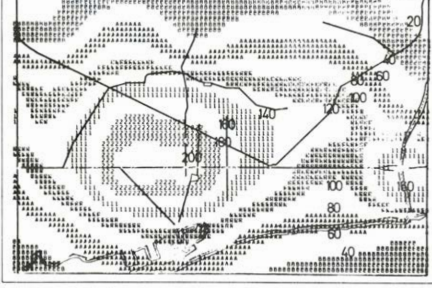


PRIMAVERA

VERANO



OTOÑO



INVIERNO

RESULTADOS PRONOSTICADOS por el ordenador para la contaminación media en Barcelona en cada estación del año (isolíneas rotuladas en microgramos por metro cúbico). Se usó el inventario de emisiones de 1970 y los valores medios estacionales de las variables meteorológicas. El modelo de difusión empleado era una modificación del de Gifford y Hanna elaborada por Roberto del Llano, entonces colaborador del autor (actualmente en el Centro de Investigación IBM de la Universidad Autónoma de Madrid). Se puede observar cómo, bajo condiciones medias, el invierno es la estación más contaminada, seguido del otoño. La predicción de contaminación para pocas horas requiere datos (decontaminación y meteorológicos) en tiempo real y exige una red de sensores mucho más densa, complicada y cara.

bles frente a la región donde se desea calcular la contaminación que produce) mientras que con frecuencia interesa calcular la contaminación causada por fuentes extensas (por ejemplo, un suburbio industrial o residencial). Todavía es utilizable en este caso la fórmula del penacho gaussiano a condición de sustituir la fuente extensa por otra puntal de intensidad adecuada y situada viento arriba de la fuente extensa a una distancia convenientemente elegida y al nivel conveniente, que puede ser el del suelo.

El modelo de penacho gaussiano no es el único, aunque sí el más sencillo y versátil. Mediante retoques adecuados puede aprovecharse para fuentes extensas, para viento en calma y para diversas condiciones de estratificación atmosférica incluyendo inversiones.

Se plantea ahora la posibilidad de extenderlo al cálculo de la contaminación producida por un conjunto de fuentes de las cuales algunas serán puntuales, otras extensas y otras difusas (es decir, no fijas ni de intensidad constante, como por ejemplo el tráfico rodado). Tal conjunto complicado de fuentes corresponde a la situación existente en una ciudad populosa, con un área más o menos fuertemente industrializada y otra predominantemente residencial y con las correspondientes vías de comunicación rápidas y lentas, urbanas e interurbanas.

En este caso es preciso comenzar por realizar un censo o inventario de las fuentes puntuales, al menos las más importantes, que especifique su situación, intensidad, clase de contaminantes que produce y horas de funcionamiento; hay después que pensar cómo tratar las fuentes pequeñas y dispersas (por ejemplo, chimeneas domésticas), que pueden englobarse en una fuente extensa por cada manzana, cuadra o bloque de casas atribuyéndole una intensidad adecuada, y finalmente, el tráfico puede simularse mediante fuentes lineales a lo largo de las calles o vías con una intensidad proporcional a la densidad del tráfico y a la cantidad de contaminantes que por término medio emite cada vehículo.

Cuando se dispone de todo ello se tiene el punto de partida para estimar la distribución de contaminantes en la ciudad considerada. Es necesario ahora aplicar a cada clase de fuente y de manera apropiada el modelo gaussiano u otro de los existentes conociendo además los valores de las variables meteorológicas pertinentes, y obtener en un número suficiente de puntos la concentración resultante, de modo que se puedan trazar isolíneas que describan su distribución espacial. Naturalmente, el número y la complicación de los cálculos requiere el uso de ordenador para obtener el producto final.

Tal procedimiento constituye lo que se llama un modelo urbano de difusión. F. A. Gifford y Steven R. Hanna, del antes citado Laboratorio de Turbulencia y Difusión Atmosférica, figuran entre los primeros en haber ideado un tal modelo que fuera prácticamente aplicable. Los refinamientos posteriores han consistido en introducir modelos de difusión más complicados que el gaussiano (sin excluir métodos estadísticos) o en tener en cuenta la topografía y otros efectos locales. Los modelos se pueden aplicar bien a valores medios (tanto de emisiones como de datos meteorológicos) como a valores instantáneos de unos y otros. En el primer caso se obtiene un modelo climatológico, cuya utilidad estriba en conocer cuál será la distribución de contaminantes en cierta época del año en la ciudad en cuestión cuando se prescribe una cierta cantidad y distribución de focos emisores. En la ilustración de la página 117, se ve el resultado obtenido para Barcelona por Roberto del Llano, entonces colaborador del autor, haciendo uso del censo de emisiones para 1970 y de una versión modificada del modelo de Gifford-Hanna.

Cuando se desea obtener valores reales de concentraciones en un instante dado es preciso usar un "modelo episódico", en el que los datos de entrada de concentraciones y meteorológicos proceden de una red automática de sensores y captadores adecuadamente distribuida que los transmite en tiempo real al ordenador el cual dispone también del inventario y el programa de cálculo. Tal sistema, naturalmente mucho más costoso, permite la toma de decisiones por parte de las autoridades responsables a la vista de lo que vaya sucediendo, y es particularmente útil en situaciones críticas o lindantes con ellas.

¿Puede decirse que con ello el problema está resuelto? Para saberlo habría que evaluar el modelo, es decir, comparar sus predicciones para ciertos valores de entrada con las correspondientes observaciones durante el mismo período. La dificultad estriba en que las mediciones, realizadas en cierto número de puntos, son representativas de su entorno inmediato mientras que los modelos proporcionan, en general, valores medios de la concentración en un área correspondiente a la dimensión de la red discreta usada para la solución numérica de las ecuaciones, que suele ser de uno a cuatro kilómetros cuadrados. Las comparaciones, por consiguiente, no pasan de ser cualitativas pero los resultados, si no más, son alentadores.

¿Hasta qué punto puede la contaminación atmosférica afectar al clima del planeta? El clima en la superficie terrestre está determinado por un complicado cuasi-equilibrio entre la energía recibida del Sol y su distribución entre la atmósfera y la superficie terrestre (sea ésta continental, oceánica o superficie de hielo). Como se dijo, la atmósfera misma apenas se calienta al ser atrave-

VARIACION DE LA CONCENTRACION de dióxido de carbono con el tiempo medida en varios lugares (en negro, vuelos de reconocimiento sobre Suecia; en gris, Point Barrow, Alaska; en rojo, Polo Sur; en azul, Mauna Loa, Hawai) desde 1958 y las estimadas por diversos investigadores basándose en datos históricos dispersos hasta 1865 y extrapolando la tendencia actual hacia el futuro según modelos especulativos que tienen en cuenta diversos factores. Esta figura se ha adaptado a partir de otra de William W. Kellogg, del Centro Nacional de Investigación Atmosférica de Boulder, estado de Colorado. Como la importancia relativa de los factores no se conoce bien ni tampoco la interacción entre ellos, las estimas de concentración del dióxido de carbono para el año 2000 discrepan ampliamente.

sada por la radiación solar porque es casi transparente para ella; lo que sí se calienta es la superficie terrestre, que a su vez emite radiación no visible de mayor longitud de onda, llamada infrarroja o térmica. Las nubes y la propia atmósfera emiten también radiación infrarroja, de la cual aproximadamente la mitad se perdería en el espacio. Pero a diferencia de lo que ocurre con la radiación solar, algunos gases atmosféricos, en particular el dióxido de carbono CO_2 y el vapor de agua, absorben una proporción substancial de la radiación infrarroja impidiendo que escape al espacio y devolviendo parte de ella a la superficie terrestre, que tiende así a calentarse. Las regiones tropicales reciben del Sol más energía que la correspondiente al balance, mientras que las regiones polares son deficitarias. La compensación se realiza gracias al transporte de calor por la atmósfera (alrededor del 90 por ciento) y por los océanos, de modo que para un período de tiempo suficientemente largo, la temperatura media de la superficie terrestre permanece aproximadamente constante en cada lugar.

La contaminación se debe fundamentalmente a la clase de actividad humana que comporta la industrialización, y los países industrializados son hoy una minoría en el mundo. Sin embargo, se conocen dos efectos sobre la atmósfera que están empezando a causar preocupación: uno es el aumento de concentración del dióxido de carbono y otro el aumento de concentración de aerosoles.

El contenido de dióxido de carbono en la atmósfera ha aumentado continuamente desde la revolución industrial; se estima que desde 1860 el aumento ha sido del 10 por ciento del total de CO_2 presente en la atmósfera natural, y entre 1958 y 1968 ha aumentado en un dos por ciento [*véase la ilustración de la página 118*]. Las cantidades no parecen alarmantes, pero si se recuerda que el dióxido de carbono es uno de los principales absorbentes de radiación infrarroja, resulta que el aumento de CO_2 en la atmósfera debe traducirse en un aumento de su temperatura media.

Por otra parte, debe considerarse el efecto del aumento de concentración de aerosoles. Anualmente se incorporan a la atmósfera entre 30 y 70 millones de toneladas de aerosoles, de los que 3 a 14 millones de toneladas son debidas a la actividad humana. En la estratosfera, los aerosoles absorben radiación ultravioleta, tendiendo a calentarla, pero el efecto es pequeño. En la troposfera absorben y difunden radiación solar. El efecto directo es también pequeño, pero al interceptar parte de la radiación que calentaría el suelo, tienden a producir un descenso de temperatura que compensaría, al menos parcialmente el calentamiento debido al CO_2. Sin embargo, no todos los especialistas están de acuerdo en que el aerosol produzca enfriamiento porque su efecto en el infrarojo no está suficientemente explorado.

Para estimar el efecto a largo plazo de estos aumentos de contaminantes no basta con la simple extrapolación. La cuestión se ha enfocado simulando el comportamiento del sistema Tierra-atmósfera con ciertas hipótesis simplificadoras que permitan formular matematicamente el problema. Cada conjunto de hipótesis da lugar a un modelo climático, pero ni la complejidad del sistema y nuestros insuficientes conocimientos del mismo ni la capacidad de los ordenadores permiten elaborar modelos climáticos completamente satisfactorios. Una y otra cosa explican que los más destacados especialistas en este campo estén en desacuerdo respecto, por ejemplo, de la temperatura media global que puede esperarse en el año 2000. Así, mientras H. H. Lamb, de la Universidad de Anglia Oriental (Reino Unido), y Reid A. Bryson, de la Universidad de Wisconsin, consideran probable un ligero enfriamiento, M. I. Budyko, del Observatorio Geofísico Principal de Leningrado, y William W. Kellogg, del Centro Nacional de Investigación Atmosférica de Boulder, Colorado, se inclinan por un calentamiento de medio a un grado Kelvin; incluso hay investigadores que consideran prematura la predicción. En otras palabras, respecto a los posibles efectos climáticos del aumento de CO_2 y aerosoles en la atmósfera, no estamos seguros ni siquiera de nuestra propia incertidumbre. Esta situación ha motivado el que algunos científicos bien conocidos, en particular Hugh W. Ellsaesser, del Laboratorio Lawrence Livermore de la Universidad de California, hayan manifestado claramente su disconformidad con el catastrofismo que transparentan tanto gran parte de los estudios sobre el tema como la posición oficial en algunos países. El presente autor considera la postura de Ellsaesser plenamente justificada. No está claro si un aumento o una disminución de medio a un grado en la temperatura global haría el planeta más o menos agradable de habitar ni los presentes modelos climáticos son capaces de predecir en forma inequívoca cómo reaccionaría el sistema Tierra-atmósfera frente a tal variación. El problema, sin embargo, es importante para el futuro de la humanidad y justifica los esfuerzos de quienes trabajan en este campo.

Juegos matemáticos

Cómo imaginarse los números imaginarios

Martin Gardner

"Los números imaginarios son vuelo maravilloso del espíritu divino, casi anfibios entre el ser y el no ser."

—LEIBNIZ

En la sección de agosto de 1977, dedicada a los números negativos, expliqué cuán largo tiempo necesitaron los matemáticos para ampliar el concepto de "número" hasta incluir entre ellos a los negativos. El mismo proceso se repitió, con desazón todavía mayor, cuando los matemáticos descubrieron la enorme utilidad de los números tan desafortunadamente bautizados como "imaginarios". Su historia es tan extraña como hermosa.

No obstante algunos tempranos pronunciamientos sobre la imposibilidad de que los números negativos admitan raíces cuadradas (debido a que los cuadrados de los números reales son positivos o cero), la historia de los números imaginarios comienza en realidad en la Europa del siglo XVI. Los matemáticos de la época, y particularmente el italiano Rafaello Bombelli, descubrieron que al resolver ecuaciones algebraicas solía ser útil admitir que los números negativos sí tenían raíces cuadradas. Dicho de otra forma, al igual que la ecuación $x + 1 = 0$ solamente podría resolverse admitiendo que x tomara el valor -1, la única forma en que podría resolverse $x^2 + 1 = 0$ sería tomando $x = \sqrt{-1}$.

La presunción, en apariencia ridícula, de que pueda existir alguna raíz cuadrada de -1 no tenía más justificación que su interés práctico: simplificaba ciertos cálculos, y, por tanto, sería lícito servirse de ella en tanto resultasen al final valores "reales". El paralelismo con las reglas de manejo de números negativos es sorprendente. Puede que, al intentar hallar cuántas vacas hay en un campo (es decir, al trabajar en el dominio de los enteros positivos), los números negativos faciliten los cálculos en algún paso, aunque desde luego la solución final tiene que venir dada por números positivos, pues no existe nada semejante a vacas "negativas".

A lo largo de los siglos XVII y XVIII los matemáticos de todo el mundo continuaron descubriendo nuevas aplicaciones de las raíces cuadradas de números negativos. Fue Leonhard Euler quien, en el siglo XVIII, introdujo el símbolo i (primera letra de la palabra latina *imaginarius*) para denotar la $\sqrt{-1}$. Una afirmación muy citada, atribuida a Euler, aseguraba que tales raíces no son nada en absoluto, nada más ni nada menos que nada; son estrictamente imaginarias, o imposibles. Los matemáticos llegaron por fin a elaborar las reglas algebraicas de manejo de los "imaginarios puros" (productos de i por números reales), y más tarde, de los que hubieron de llamarse números complejos (sumas de números reales y números imaginarios puros).

Los números complejos son de la forma $a + bi$, siendo a y b números reales cualesquiera. (En este caso, el signo "más" no denota adición en el sentido familiar; su principal función es separar las partes real a y puramente imaginaria bi del número complejo.) Dicho de otra forma, si a es igual a 0 y b no lo es, el número complejo es imaginario puro, de la forma bi. Cuando b sea igual a 0, el término bi desaparece, quedando entonces el número real a. Así pues, los números complejos contienen como subconjuntos a todos los reales y a todos los imaginarios puros, al igual que los números reales contienen todos los enteros, los fraccionarios y los irracionales. Con terminología moderna, los números complejos forman la estructura matemática llamada cuerpo, cuyos elementos obedecen todas las leyes de la aritmética ordinaria. El cuerpo de los números complejos es también cerrado respecto de la adición, sustracción, multiplicación y división por números distintos de cero; es decir, aplicando estas operaciones a cualquier par de complejos se generará siempre otro elemento del cuerpo. En cierto sentido, el descubrimiento del cuerpo complejo culmina el álgebra tradicional, pues en él es posible resolver completamente toda ecuación algebraica ordinaria. El cuerpo resultó también ser cerrado con respecto a las operaciones utilizadas en análisis maté-

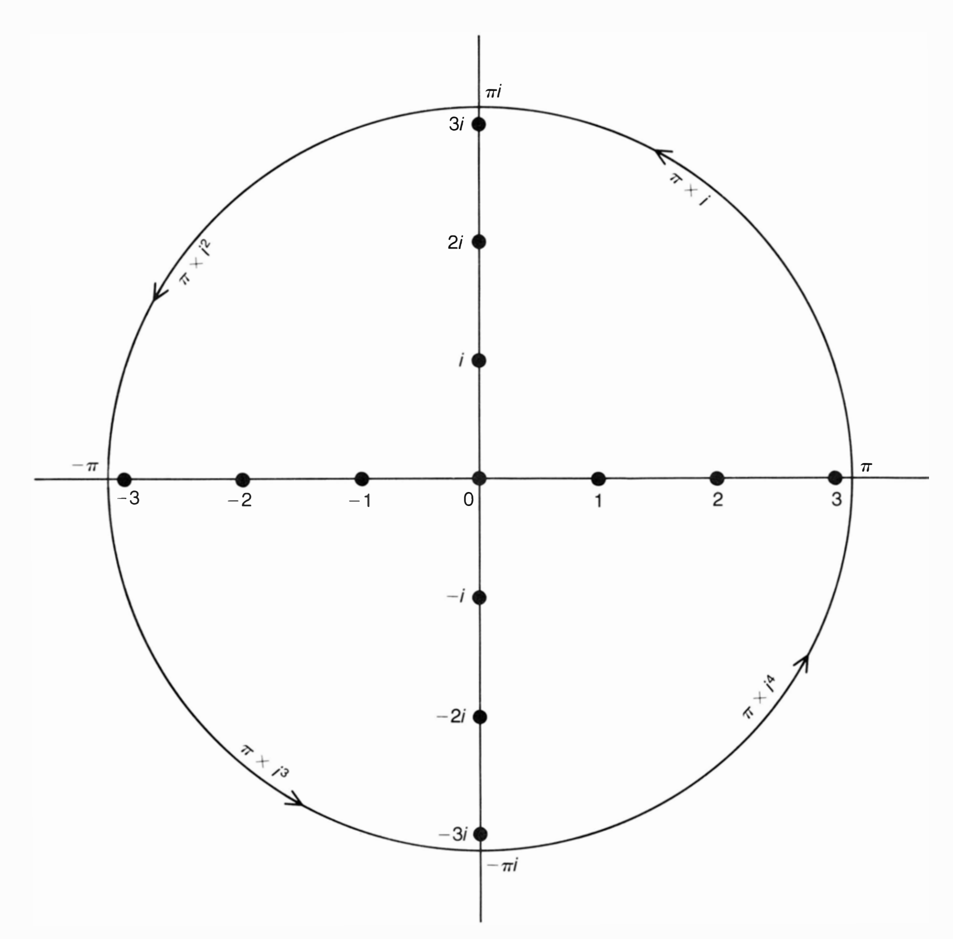
La correspondencia entre números complejos y puntos del plano

mático, y tal descubrimiento originó el vasto edificio matemático dedicado a las funciones de variable compleja.

Muchos de los progresos de la física moderna hubiesen sido imposibles de no ampliarse el álgebra a los números complejos. La primera aplicación científicamente importante de los complejos se debe a Charles Proteus Steinmetz, quien los encontraba insustituibles para efectuar eficientemente los cálculos de corriente alterna. En la actualidad, ningún ingeniero eléctrico podría pasarse sin ellos, y lo mismo vale para los físicos que trabajan en aerodinámica o en dinámica de fluidos. Los números complejos tienen también papel fundamental en teoría de relatividad (donde la simetría espacio-tiempo se consigue mediante la estratagema de considerar reales las tres dimensiones espaciales y como imaginaria la dimensión temporal), en mecánica cuántica y en otras muchas ramas de la física moderna.

Subsisten todavía recelos al afirmar que *i* es un número, por lo que no es extraño tropezarse con físicos, filósofos, e incluso algún que otro matemático que sostienen que *i* no es en realidad un número, sino tan sólo un símbolo que denota una operación que explicaré más adelante. Nadie ha sabido deshacer tal sofisma con mayor soltura que Alfred North Whitehead. En el capítulo sobre números complejos de su *Introduction to Mathematics* escribió:

"A estas alturas puede ser conveniente hacer notar que existen ciertos intelectos cuya preocupación constante –para sí y para los demás– es la licitud del uso de ciertos términos técnicos. ¿Pueden, con propiedad, ser llamados números los números inconmensurables? ¿Son verdaderamente números los enteros negativos? ¿Son los números imaginarios producto de la imaginación? Más aún, ¿son siquiera números? Todas estas cuestiones son fútiles. En efecto, nunca se habrá comprendido suficientemente que, en ciencia, los términos técnicos son nombres asignados arbitrariamente, como lo son los nombres de pila de los recién nacidos. Carece de sentido preguntarse si tales nombres son o no correctos. Su elección podrá haber sido sensata o insensata, pues a veces puede conseguirse que sean fáciles de recordar, o que susciten ideas de importancia e interés. Pero el principio esencial que debe tenerse presente fue enunciado con toda claridad en el País de las Maravillas, cuando Humpty Dumpty dice a Alicia, respecto de su forma de hablar y usar las palabras: 'Las pago extra y hago que signifiquen lo que yo quiera'. No nos preocuparemos pues de saber si los números imaginarios son imaginarios, o de si son números, sino que tomaremos tal expresión como el nombre arbitrariamente dado a una cierta idea matemática, que nos disponemos a explicar".



Productos de π por i, i^2, i^3 e i^4

El comportamiento de los números complejos es tan semejante al de los números ordinarios en lo que respecta a la adición, sustracción, multiplicación y división (de acuerdo con las reglas del cuerpo complejo) que la mayoría de los matemáticos no dudan ya en llamarlos números y en atribuirles tanta "realidad" como a los números negativos. Ni aun los números naturales pasan de ser símbolos manipulados según las reglas de un sistema deductivo. La única razón para figurárnoslos más "reales" que otros tipos de números se debe a que sus aplicaciones están tan próximas a experiencias prácticas –contar dedos, vacas, personas– que nos olvidamos de que únicamente son reales los dedos, las vacas o las personas, y no los símbolos a que recurrimos para contarlos. En el dominio de la matemática pura, tan "real" es *i* como puede serlo el 2. Si queremos, podemos pensar que 2 no es más que un operador: un símbolo que señala la duplicación de 1.

Sin embargo, casi todo el mundo está tan habituado a trabajar con los llamados números reales que al descubrir que existe una interpretación geométrica sencilla para los complejos sienten gran alivio. Esta interpretación, con la que tan fácil es "ver" qué son los complejos, consiste en identificar cada número complejo con un punto del plano cartesiano. El primero en establecer tan ingeniosa relación fue Caspar Wessel, agrimensor noruego y matemático autodidacta, quien dio conferencias sobre este tema en 1797. Unos cuantos años más tarde la idea fue redescubierta por un contable suizo, Jean-Robert Argand (quien la expuso en un libro publicado en 1806), e, independientemente, por el gran matemático alemán Carl Friedrich Gauss.

Como se ve en la ilustración de la página opuesta, la esencia de la idea estriba en considerar el eje horizontal del plano cartesiano como la recta numérica real, y el eje vertical como recta de puntos correspondientes a los números puramente imaginarios. Dicho de otra forma, se establecen correspondencias biunívocas entre los números reales y los puntos del eje del x y entre los números imaginarios puros y los puntos del eje y. Como ya he señalado, estos dos conjuntos pueden considerarse como subconjuntos de números complejos, y ahora, los restantes números complejos pueden ponerse en correspondencia

cia biunívoca con los restantes puntos del plano. Para obtener las coordenadas de un punto asociado a un número complejo basta medir su parte real sobre el eje real y la parte imaginaria, sobre el eje imaginario. En la ilustración se muestran los puntos correspondientes a cuatro números complejos.

Con esta interpretación de los complejos es posible olvidar completamente la desazonadora noción de que i es la raíz cuadrada de -1 (que evidente no es, en el sentido ordinario de la operación de tomar raíces cuadradas). En adelante, un número complejo puede considerarse como un par ordenado de números reales, simplemente; el primero de ellos tomado sobre el eje real y el segundo, sobre el eje imaginario. Con otras palabras, definiendo convenientemente las operaciones aritméticas de composición de estos pares es posible construir un álgebra de pares ordenados de números reales equivalente al álgebra de los números complejos. La obscura frase "raíz cuadrada de un número negativo" no aparece ya en ningún lugar de esta nueva álgebra, aunque, desde luego, la misma idea está presente bajo lenguaje y notación distintas. Si este álgebra de pares ordenados hubiese sido descubierta antes que los complejos, quizá hoy no habría quien recordase los números imaginarios ni se preguntase si existen verdaderamente o no.

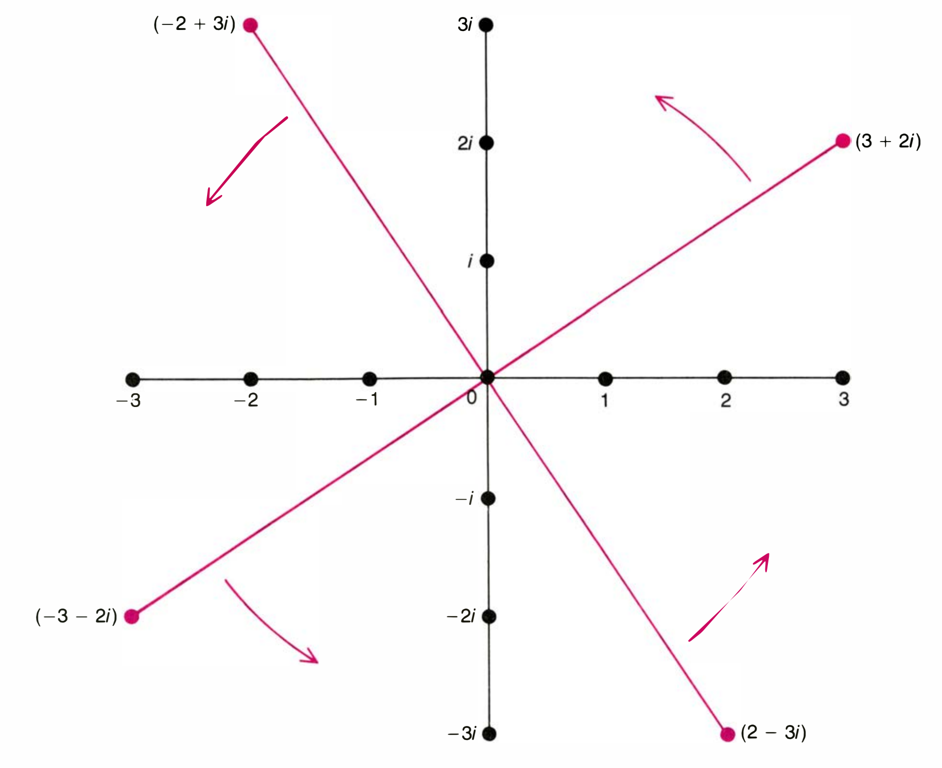
Tras el descubrimiento de esta interpretación geométrica de los números complejos, los matemáticos se preguntaron inmediatamente si sería factible generalizar la idea fundamental a tres dimensiones, es decir, a los puntos del espacio. Desdichadamente, la respuesta es negativa a menos que se modifiquen radicalmente las leyes de la aritmética. Fue el matemático irlandés William Rowan Hamilton quien consiguió los primeros progresos en la teoría de "números hipercomplejos", inventando los cuaternios: números compuestos de cuatro partes, que combinan una parte real con tres imaginarias. La clave de su manejo se encuentra en que ya no obedecen la ley conmutativa, a saber, la regla que estipula que para todo par de números a y b, ab es igual que ba.

La idea de sacrificar la ley conmutativa se le ocurrió a Hamilton un atardecer, mientras paseaba con su esposa a lo largo del Royal Canal de Dublin. Tan impresionado quedó, que se paró a garrapatear la fórmula fundamental en una de las piedras de Brougham Bridge. Ya en tiempos de Hamilton las inclemencias atmosféricas borraron la inscripción original; hoy el puente es únicamente conocido por ser el que cruza Brombridge Street. No obstante, hay en la piedra una placa que conmemora el gran suceso, y en 1943, un siglo después de la revelación de Hamilton, Irlanda emitió un sello de correos en su honor. Los cuaternios no forman cuerpo (su estructura se llama anillo con división), pero el álgebra de cuaternios es equivalente a un álgebra de cuadrupletas ordenadas y es hoy frecuentemente aplicada como parte de la teoría de vectores tridimensionales. El descubrimiento del álgebra de cuaternios señaló el comienzo del álgebra abstracta moderna, donde pueden definirse toda clase de "números" mucho más raros y sorprendentes que los números complejos.

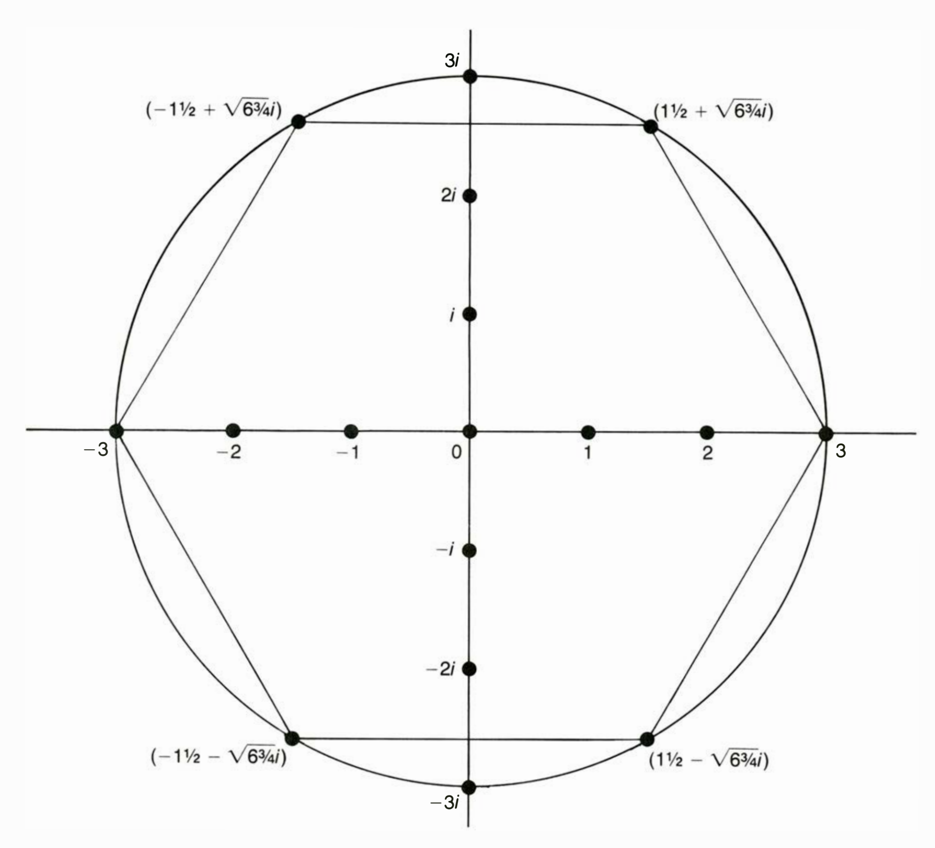
En vista de la correspondencia entre números complejos y puntos del plano cartesiano, cuando el plano se maneja de esta forma se llama "plano complejo". (También es llamado plano z, por la notación $z = a + bi$ utilizada para denotar complejos, y, también, diagrama de Argand, pues durante muchos decenios nadie tuvo noticia de que Wessel se anticipara en el descubrimiento.) No entraré aquí en detalles sobre la adición, sustracción, multiplicación y división de números complejos mediante diagramas geométricos trazados en el plano complejo. Los lectores que no conozcan todavía las reglas que gobiernan estas operaciones pueden consultarlas en cualquier texto de álgebra elemental que trate el tema. No obstante, sí es necesario exponer brevemente el producto por i para poder presentar un elegante teorema relativo a las raíces de los números.

Para multiplicar un número del plano complejo por i se toma su radio vector (el segmento rectilíneo que va desde el origen del plano hasta el punto) y se gira 90 grados en sentido contrario a las agujas del reloj; el nuevo extremo del vector corresponde al producto del número por i. En este sentido, i puede ser considerado como un operador. Para comprender mejor esta idea, observemos qué ocurre al elevar i a distintas potencias. Evidentemente, i elevado a la primera potencia es igual a i, y es fácil ver que i^2 es igual a -1, que i^3 es igual a $-i$, y que i^4 es igual a 1. Este ciclo de cuatro etapas se repite indefinidamente: i^5 es igual a i, i^6 es -1, i^7 es $-i$, i^8 es igual a 1, y así sucesivamente. Todas las potencias pares de i son iguales a 1 o a -1, y todas las potencias impares, iguales a i o a $-i$.

La ilustración de la página anterior muestra de qué forma se aplican estas observaciones al producto de un número (en este caso, pi, o sea, π) por i. Una vez localizado el punto del lado positivo del eje x correspondiente a π, se le hace girar 90 grados en sentido antihorario a lo largo de una circunferencia de radio π cuyo centro es el origen del plano. Una flecha indica que el punto resultante de este proceso es un imaginario puro, πi, situado en la parte superior del eje y. Multiplicar π por i^2 es pues equivalente a multiplicar dos veces consecutivas por i: el punto correspondiente a π es desplazado 180 grados a lo largo de la circunferencia y acaba en el



Productos del número complejo $3 + 2i$ por i, i^2, i^3 e i^4



Las seis raíces sextas de 729

punto $-\pi$ del eje x, o sea, de la recta numérica real. Análogamente, al multiplicar π por i^3 se efectúa un giro de 270 grados, que termina en el punto $-\pi i$ sobre la parte inferior del eje y; multiplicar por i^4 es lo mismo que multiplicar por 1, con lo que volvemos a π. Podemos proseguir de la misma forma con todas las potencias superiores de i. Cada sucesiva potencia nos traslada otro cuarto de vuelta en sentido antihorario alrededor de la circunferencia.

La operación recíproca de multiplicar por i es dividir entre i, a saber, girar 90 grados en sentido horario, alrededor del origen del plano. O sea, dado un número complejo cualquiera, se traza el segmento radio vector desde el origen hasta el punto que represente al número. Entonces, para multiplicar el número por i se gira el vector 90 grados en sentido antihorario [*véase la ilustración de la página anterior*], mientras que para dividirlo por i el vector se gira 90 grados hacia el otro lado. (Un amigo me decía, en broma, que el producto de i por el infinito es igual a 8, porque la multiplicación de i por el signo de infinito lo pondría vertical.)

Con esta interpretación del producto resulta que todo número distinto de cero, sea real o complejo, tiene exactamente n raíces n-ésimas, contadas las que sean complejas. Es decir, cada número tiene dos raíces cuadradas, tres raíces cúbicas, cuatro raíces de cuarto orden, cinco raíces quintas, etcétera. Se deduce que toda ecuación cúbica tiene tres soluciones, toda ecuación de cuarto grado tiene cuatro soluciones, y así sucesivamente; además, al representar gráficamente en el plano complejo las raíces de números individuales queda de relieve una inesperada y preciosa propiedad. Los n puntos correspondientes a las n raíces n-ésimas yacen todos sobre una circunferencia cuyo centro es el origen del plano, y equidistantes sobre ella. Esto es, tales puntos definen los vértices de un polígono regular de n lados. Por ejemplo, la ilustración de esta página muestra la localización de las seis raíces sextas de 729. Si, como ocurre en este caso, el número es real y tiene un número par de raíces, dos de los vértices del polígono se encontrarán sobre el eje real. Si el número es real, y el número de raíces es impar, solamente un vértice del polígono yacerá sobre el eje real.

Además de ser fundamental en física moderna, el cuerpo de los números complejos ha enriquecido la matemática pura con multitud de terroríficos teoremas. Es necesario tener presente que los números complejos, aunque contienen como subconjunto a los reales, se diferencian de ellos en inesperadas formas. Por ejemplo, no puede decirse de un complejo que es positivo o negativo; estas propiedades sólo son aplicables a

los reales y a los imaginarios puros. Igualmente carece de sentido afirmar que un complejo es mayor o menor que otro.

Ya desde antes de Euler se sabía que el producto de dos imaginarios puros es un número real; pero fue Euler quien por primera vez hizo ver que i^i es también real. Resulta ser un número irracional cuyo desarrollo decimal nos daría 0,2078795763... En realidad, este número es solamente uno de los infinitos valores, todos reales, de i^i. El conjunto de todos ellos está dado por la fórmula $e^{-\pi/2 \pm 2k\pi}$ siendo k un entero cualquiera, por lo que cuando k es igual a 0, se obtiene el valor principal dado anteriormente. También la i-ésima raíz de i es un número real, cuyo valor principal es $e^{\pi/2}$, que viene a ser aproximadamente igual a 4,8104773809...

Hay otras muchas fórmulas donde i entra en relación con los dos irracionales más conocidos, e (la base de los logaritmos naturales) y π. La fórmula más famosa, debida a Euler, aunque basada en otro descubrimiento anterior, es $e^{i\pi} + 1 = 0$, de la que Edward Kasner y James Newman dicen en su libro *Mathematics and the Imagination* que es "elegante, concisa y llena de significado". También citan una observación de Benjamin Peirce, matemático en Harvard y padre que fue de Charles Sanders Peirce, referente a esta fórmula: "Señores", dijo tras escribirla en la pizarra, "estamos seguros de que es cierta, (aunque) sea absolutamente paradójica; no podemos entenderla, no sabemos qué significa, pero la hemos demostrado, y por consiguiente sabemos que tiene que ser verdadera".

Bueno, en realidad la fórmula no carece tan completamente de sentido. Escrita en la forma $e^{i\pi} = -1$ podemos representarla gráficamente en el plano z como límite de la sucesión infinita $1 + \pi i + (\pi i)^2/2! + (\pi i)^3/3! + (\pi i)^4/4!$.... (El signo de exclamación representa la factorial: $n!$ es igual al producto $1 \times 2 \times 3 \times ... \times n$.). Los términos de esta sucesión se representan gráficamente como una sucesión infinita de puntos situados sobre una espiral de líneas rectas de sentido antihorario que "estrangulan" el punto -1 del eje real.

George Gamow, buscando disipar el misterio de los números complejos, preparó el siguiente problema. Un antiguo pergamino daba al describir la situación de un tesoro pirata enterrado en una isla desierta estas instrucciones: En la isla hay tan sólo dos árboles y los restos de una horca. A partir de la horca se cuentan los pasos necesarios para llegar en línea recta hasta el árbol A. Una vez allí se giran 90 grados a la izquierda y se camina al frente el mismo número de pasos. Al cabo, se clava en el suelo una estaca. Ahora hay que volver a la horca y contar los pasos, en línea recta desde ella hasta el árbol B. Cuando se llega al árbol se giran 90 grados hacia la derecha y se camina de frente este número de pasos. Se clava otra estaca en el punto de detención. Cavando en el punto situado exactamente a medio camino entre las estacas se encontrará el tesoro.

Un joven aventurero descubrió el pergamino que contenía estas instrucciones. Fletó un barco y navegó hasta la isla. No tuvo dificultad en encontrar los árboles, pero la horca había desaparecido, y el tiempo, borrado toda huella de su emplazamiento. Desconociendo el lugar donde estuvo situada la horca, el joven no supo ver el modo de encontrar el tesoro, y tuvo que regresar con las manos vacías. Gamow señala que si este joven estuviese familiarizado con las técnicas de manipulación de números en el plano complejo habría hallado el tesoro con gran facilidad. Los lectores que conozcan las reglas básicas de representación gráfica de complejos deberían poder resolver el problema. Daré la solución el mes que viene.

El primer problema del mes pasado consistía en hallar el mínimo número entero expresable de dos formas distintas como suma de dos cuadrados. Tal número es 50, que es igual a $1^2 + 7^2$ y también a $5^2 + 5^2$. Si se permite utilizar el cuadrado de 0 la solución es 25, que es igual a $5^2 + 0^2$ y a $3^2 + 4^2$. Si ambos cuadrados han de ser no nulos y distintos entre sí, la solución es 65, que es igual a $8^2 + 1^2$ y a $7^2 + 4^2$.

El segundo problema era determinar si *MU* es o no un teorema en el sistema formal *M*, *I* y *U*. Puede verse una demostración sencilla de por qué *MU* no es un teorema en las páginas 260 y 261 del libro que recensé el mes pasado: *Gödel, Escher, Bach: an Eternal Golden Braid*, cuyo autor es Douglas R. Hofstadter.

Muchos lectores han generalizado el problema de la mesa giratoria propuesto el pasado abril. Lyle Ramshaw, Richard A. Litherland y el equipo de Stephen Willard y Ted Lewis fueron los primeros en demostrar que dada una mesa con n pozos y un jugador con k manos, es posible conseguir que el timbre llegue a sonar en un número finito de pasos si y solamente si k es mayor o igual que $n(p-1)/p$, siendo p el máximo divisor primo de n. Todavía no se ha encontrado una fórmula no recursiva que dé el número mínimo de pasos necesarios para ello.

Taller y laboratorio

Cómo hacer campanas y cortinas de agua con cuchillos, cucharas y otros utensilios

Jearl Walker

Uno de los pocos placeres que encuentro en el enojoso trabajo de lavar los platos es hacer finas películas de agua, que en algunos sitios se les ha dado en llamar campanas o cortinas de agua. Si mantenemos la parte plana de un cuchillo perpendicular a la corriente de agua, podremos lanzar hacia fuera, por uno y otro lado de la hoja, delgadas cortinas de agua con sólo ajustar el caudal de flujo y la distancia entre cuchillo y grifo. En determinadas condiciones, las cortinas de agua se pliegan hacia atrás sobre sí mismas, formando unas hermosas superficies curvadas.

Con cucharas y otros cubiertos planos o ligeramente curvos, pueden crearse también cortinas e incluso otras figuras. Cualquier pequeño obstáculo circular puede utilizarse para lanzar hacia fuera cortinas que luego se pliegan sobre sí mismas originando formas simétricas abombadas: las campanas de agua. A veces, las cortinas de agua de ciertas fuentes son tan grandes que bien pudiera llamárselas esculturas de agua. Aunque también influyen otros factores en la forma que adopten las cortinas, se debe sobre todo a la tensión superficial el que se mantengan unidas contra las fuerzas desintegradoras. Tensión superficial que, además, las pliega sobre sí mismas para formar las campanas.

En mi opinión, el procedimiento más sencillo para lograr esas campanas es el que se sirve del tapón de rosca de una botella de bebida no alcohólica. La superficie plana de la parte superior suele tener el tamaño correcto para conseguir una buena campana; puede sujetarse fácilmente introduciendo dos dedos por el fondo del tapón. Para no destruir la campana con la mano, hay que colocar el tapón a la distancia adecuada del grifo introduciéndolo por abajo. Las cortinas de agua se pliegan hacia atrás sobre el antebrazo formando grandes campanas.

Campana de agua hecha con un tapón de botella

Pueden formarse cortinas a partir de corrientes de agua proyectadas hacia arriba, como sucede en muchas fuentes públicas. Un pequeño obstáculo situado en una corriente puede lanzar la cortina de agua hacia los lados, hacia arriba o hacia abajo según la geometría del obstáculo. Podemos crear una campana de agua con una corriente hacia arriba o con una corriente hacia abajo, siempre que se ajuste la velocidad a la que el agua choca contra el obstáculo, si bien la forma de la campana quizá no sea siempre la misma. Si el obstáculo tiene una superficie cóncava, las cortinas de agua se dirigen hacia abajo originando una campana de casi la misma configuración que muestran las que surgen instalando el objeto debajo del grifo. Si el obstáculo tiene su superficie convexa, las cortinas de agua se dirigen inicialmente hacia arriba, luego se curvan y caen hacia abajo. La campana presentará una parte superior más aplastada que las habituales formas redondeadas por arriba observadas en el fregadero.

Las cortinas de agua han venido siendo objeto de atención, por lo menos, desde 1833, en que aparece el primer estudio firmado por Félix Savart. En 1935 se conocía ya la mayoría de los principales determinantes de la forma de las cortinas y campanas, aunque tal vez no en detalle debido a la dificultad que entrañan sus cálculos matemáticos. En los últimos veinte años, el interés por las cortinas de agua ha crecido, instado por la necesidad de controlar las capas de fluido surgida en algunos campos, como por ejemplo el de la pintura con spray. El proceso por el cual una delgada película de líquido se rompe formando gotitas es de importancia particular. El radio máximo de una cortina está básicamente determinado por la tensión superficial del agua. La cortina se desvanece al alcanzar un radio para el que la tensión superficial no es capaz de mantener la película de líquido unida, a pesar del empuje hacia fuera que induce el movimiento del agua, es decir, a pesar de la fuerza de inercia del líquido en su flujo radial desde el punto de impacto con el obstáculo encontrado en la corriente.

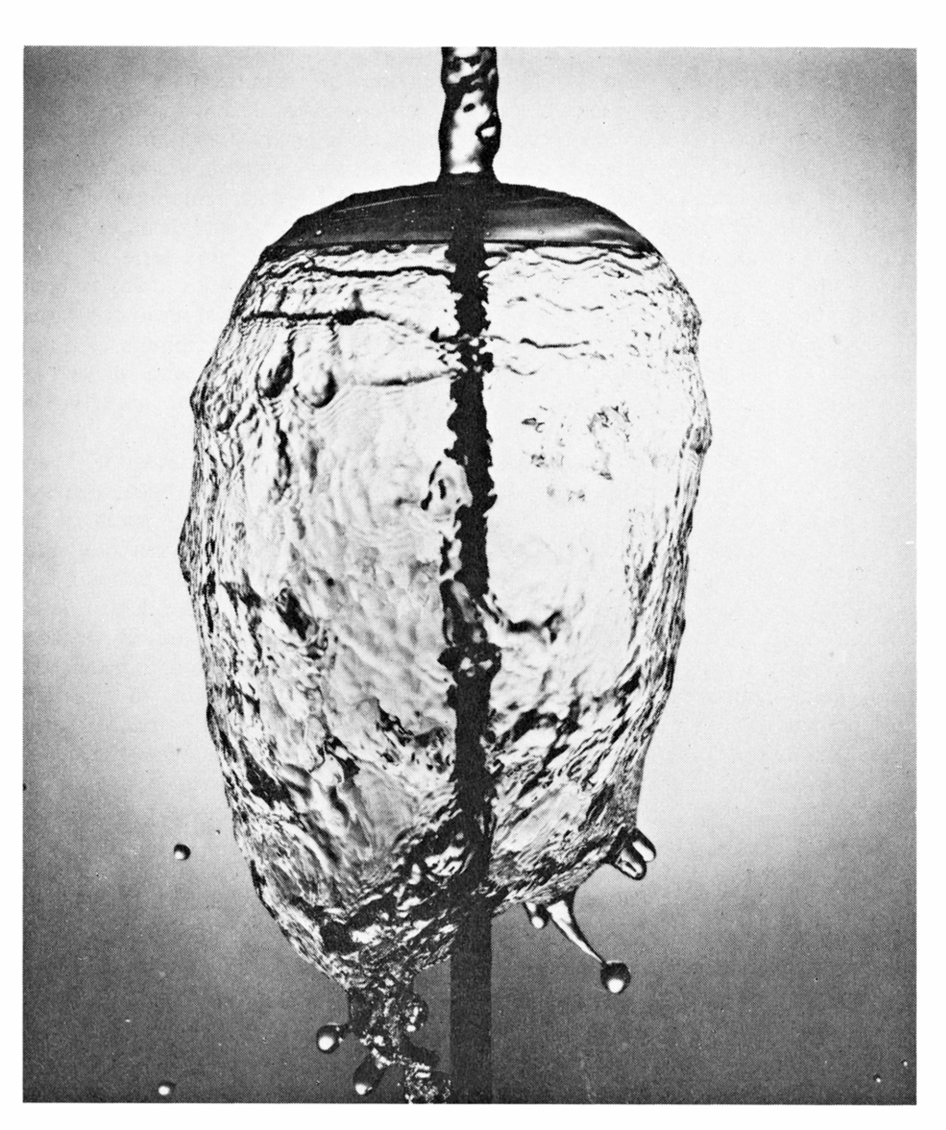
Además del tapón de botella, los objetos que mejor me han servido de obstáculos en mis trabajos con las campanas de agua han sido el fondo plano de un recipiente de plástico de 20 milímetros y un cristal de reloj de unos seis centímetros de diámetro. El recipiente era lo suficientemente grande como para poder sujetarlo con los dedos sin que el brazo entorpeciera el desarrollo de las cortinas de agua. Por el contrario, con el cristal de reloj, si bien

servía como superficie curvada de experimentación, los dedos obstruían siempre las cortinas de agua.

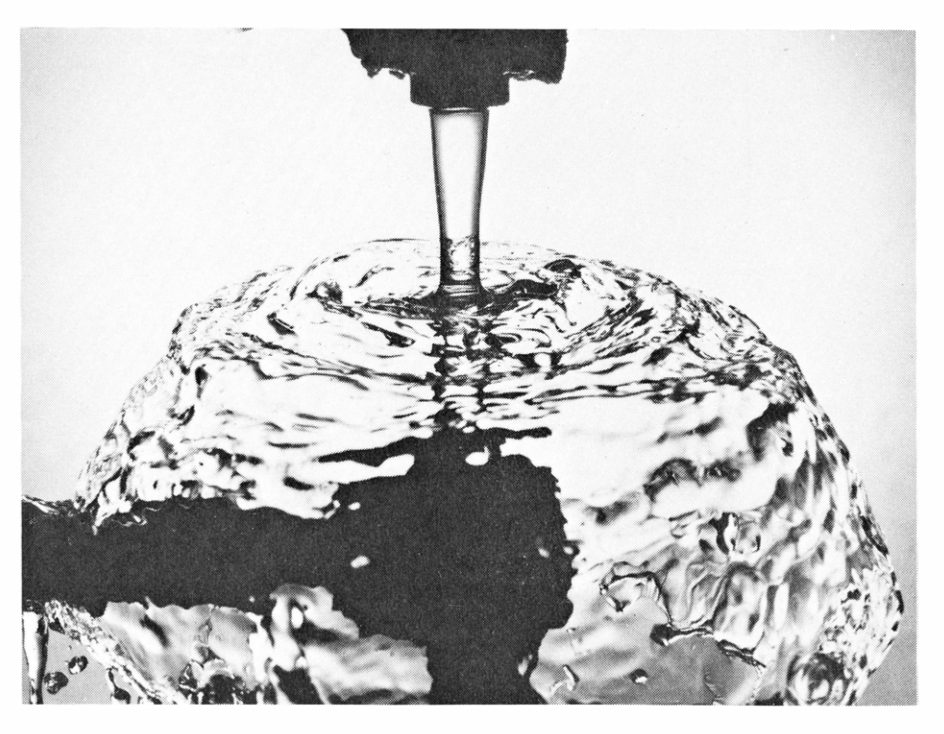
Me entretuve primero con una corriente normal de agua descendente, creando grandes cortinas y campanas de agua que, en no pocas ocasiones, terminaban en mi camisa y en el suelo de la cocina. Cuando el chorro era bastante intenso, las cortinas formadas con el recipiente derivaban hacia los lados con una cantidad de movimiento tal que alcanzaban el radio de desintegración antes de que pudieran curvarse. Con menor velocidad de impacto, las cortinas se curvaban antes de desintegrarse. A una velocidad alta obtuve cortinas casi horizontales, mientras que a velocidades más bajas conseguía campanas casi perfectas. El cristal del reloj no me permitió tales filigranas. El lado cóncavo producía una breve cortina de agua que duraba muy poco debido a la turbulencia. El lado convexo enviaba el agua inicialmente hacia abajo, posibilitando cortinas mejores.

Para obtener una corriente hacia arriba, incorporé un trozo de manguera al grifo. Levanté el extremo libre de la misma y lo fijé con una abrazadera unida a un anillo soporte. Esta disposición tenía la ventaja de producir las cortinas de agua cerca de la taza del fregadero, lo que disminuía la probabilidad de que me mojara y salpicase la cocina. El recipiente formaba grandes cortinas y campanas de agua, a tenor de cómo ajustaba la velocidad del flujo y a qué distancia del extremo abierto de la manguera estaba el fondo del recipiente. El lado cóncavo del cristal del reloj daba resultados similares. El lado convexo enviaba la cortina hacia arriba. Ajustando la velocidad del flujo, podía variarse la forma de las cortinas, así como hacer que se curvaran por la parte de arriba hasta alcanzar una altura máxima a partir de la cual comenzaban a caer. El cambio de dirección podía ser suave y gradual o bien cortante y radical. Cuando interpuse un tenedor en una corriente ascendente, se formaron varias cortinas de agua, una por cada espacio entre dientes, orientadas cada una de ellas en una dirección distinta. Además, la porción de flujo que chocaba contra la sección cóncava del tenedor producía parte de una campana de agua. El efecto combinado resultaba hermoso y fascinante.

En 1935, Frank L. Hopwood presentó una buena descripción de las campanas de agua, donde estudiaba las diferentes formas según las distintas velocidades de flujo. Hopwood construyó sus campanas haciendo pasar el agua a través de un



Campana creada con un cristal de reloj



Cortinas formadas por la colisión coaxial de dos chorros

tubo de latón de unos 10 centímetros de largo y unos 2,1 centímetros de diámetro exterior. Coaxialmente con el tubo, había una varilla que sobresalía por la cabeza del mismo, llevando incorporado en esa zona un pequeño cilindro o una tuerca. El agua subía por entre la varilla y las paredes del tubo y salía por la estrecha abertura existente entre el extremo superior del tubo y el obstáculo situado en la varilla. Enroscando o desenroscando el cilindro o la tuerca a lo largo de la varilla, Hopwood podía ajustar el tamaño de la abertura.

El tubo estaba montado verticalmente en un gran recipiente que recogía el agua que caía. Una tubería servía de rebosadero para evacuar el agua, antes de rebosar, a un fregadero próximo. Controlaba la corriente de agua con una llave de paso. Con este montaje, las campanas de agua de Hopwood se unían con la superficie del agua del recipiente produciendo campanas cerradas.

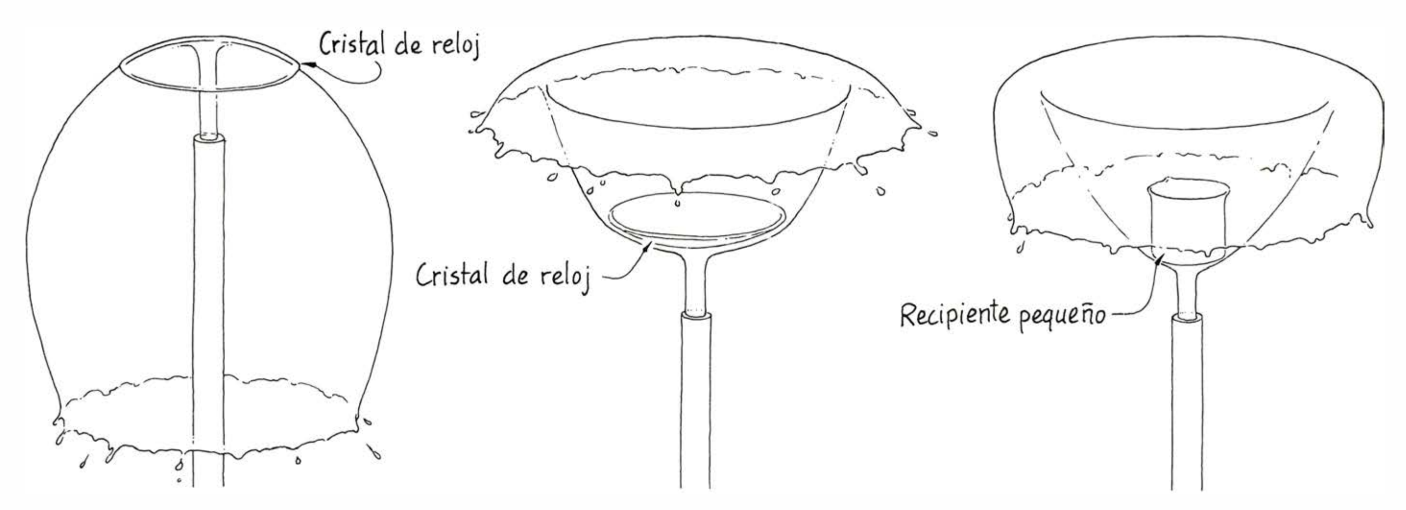
Hopwood obtuvo campanas con la bóveda redondeada y con la bóveda plana, según ajustaba la altura del obstáculo sobre el nivel del agua del recipiente, la anchura de la abertura y el caudal de agua. Algunas de las campanas obtenidas con la parte superior plana tenían los lados curvados hacia el centro, mientras que los de otras eran rectos, aunque inclinados. En ciertos casos, el borde de la zona del cambio de dirección se distinguía con nitidez. A veces el borde formaba un montículo y Hopwood observó pequeñas gotas que salían despedidas por encima del borde. Cuando construía una campana grande, y luego comenzaba a reducir lentamente el flujo de agua, observaba que dicha campana iba generando toda una serie de estas figuras estables (con los lados o bien rectos o bien cóncavos) junto con otras inestables conforme disminuía el flujo. Una vez que éste llegara a ser lo suficientemente pequeño, adquiría la forma de un cilindro vertical.

Cuando Hopwood aumentó el tamaño de la campana de agua, penetró con uno de sus dedos la pared de la campana y se encontró con que ésta dobló inmediatamente su tamaño. Al parecer, la presión del aire dentro de la campana era menor que la presión atmosférica existente fuera. Si penetraba la pared mientras hacía decrecer el caudal de agua y el tamaño de la campana, resultaba una ligera reducción del tamaño de la misma, lo que implicaba que las condiciones de presión eran diferentes.

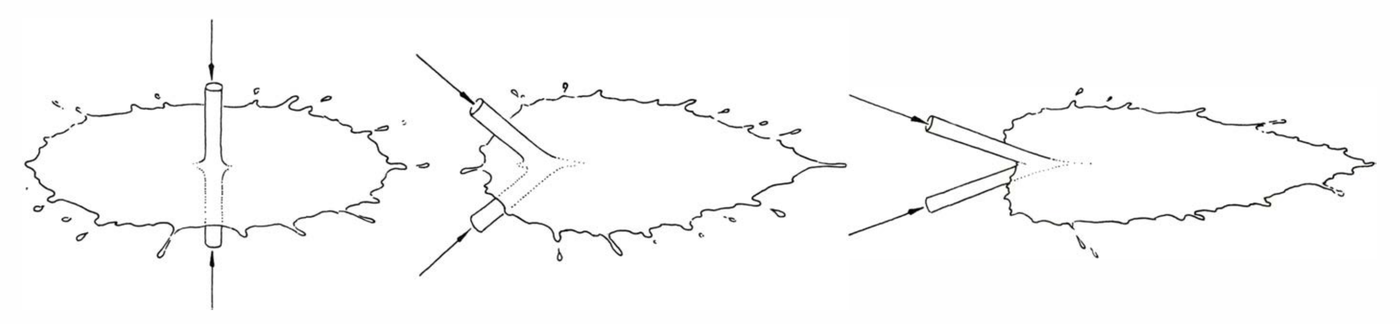
Para proseguir sus investigaciones sobre la presión del aire dentro de la campana, Hopwood inyectó poco a poco burbujas en su interior, introduciéndolas a través del agua del recipiente hasta alcanzar el centro de la campana. Si la campana era al principio redonda y ligeramente curvada, la presión extra dentro de ella incrementaba el diámetro de la base; al tiempo que disminuía el diámetro de la parte superior. Se empezaron a curvar los lados hacia dentro, comenzaba a formarse un borde en la parte superior y, a veces, aparecía un pliegue o zanja a lo largo del perímetro cerca de la base. Estos cambios continuaban hasta que el exceso de presión en el interior era tal que hacía salir el aire por la abertura de la base; entonces, la campana se encogía hasta volver a tener su forma original.

Se pueden conseguir resultados similares introduciendo una pajita flexible a través del agua del recipiente hasta el interior de la campana y soplando luego poco a poco por ella. Podría sustituirse dicha pajita por un tubito de goma flexible con una perilla en el extremo. Con uno u otro sistema pueden investigarse también los cambios que se producen al extraer aire del interior de la campana. ¿Acaso una campana con la bóveda redondeada se deshincharía de un modo regular, sin mostrar ningún otro cambio en su forma?

Hopwood sugiere que se pueden conseguir dos campanas, una dentro de la otra, con dos tubos abiertos colocados sobre un mismo eje, pero con suministro



Ejemplos de cortinas de agua formadas con una corriente invertida



Formas de cortinas generadas por dos corrientes de agua en colisión

de agua independientes. Yo mismo hice dos campanas de agua incompletas de forma más sencilla empleando el cristal curvo del reloj. Conecté el tubo del grifo a otro tubo en forma de *Y*. Uní un brazo de la *Y* a una tubería que mantuve vertical sujetándola a un anillo soporte con una abrazadera; conecté el otro a otra tubería, que monté de forma similar sobre el anillo soporte, pero en una posición superior y mirando hacia abajo. Coloqué unas abrazaderas en cada una de las mangueras, de manera que pudiera controlar la cantidad de agua que circulaba por cada una de ellas. Cuando abrí el paso del agua, introduje el cristal del reloj en la corriente. Podía ajustar el flujo de agua en cada tubería, la distancia entre orificios y la distancia entre el obstáculo y cada uno de los orificios.

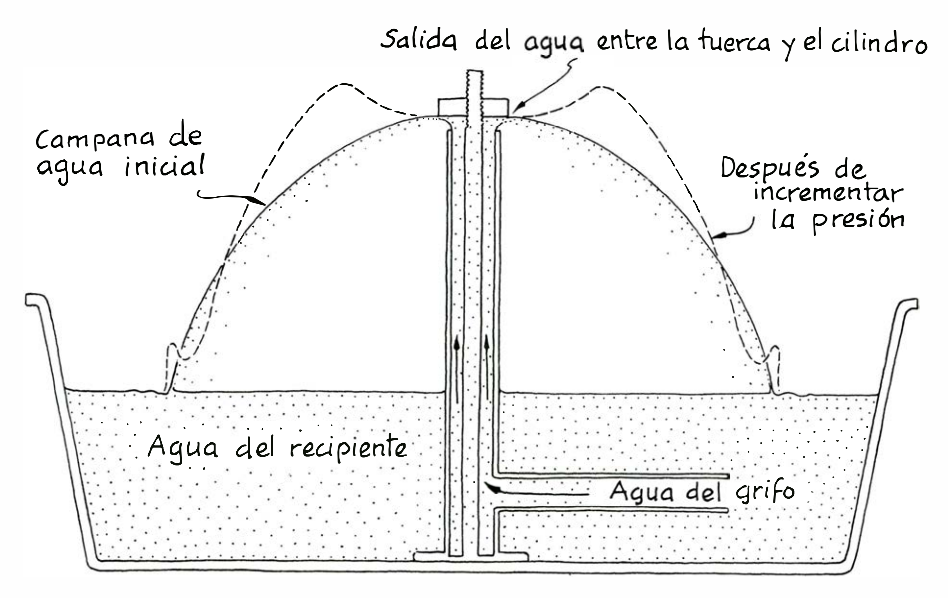
Tanto la superficie superior como la inferior del cristal del reloj producían cortinas de agua. Normalmente, las cortinas se mezclaban en el borde del cristal. Con un poco de paciencia, logré que la parte superior cóncava produjera una cortina de agua por encima que no afectaba ni era afectada por la cortina que la parte inferior convexa enviaba hacia abajo. Excepto por la región por donde sujetaba el cristal, obtuve dos campanas de agua independientes una de la otra.

Al inclinar el cristal del reloj, las dos cortinas de agua entraron en contacto y se fusionaron. Separarlas de nuevo me costó trabajo, pues la tensión superficial tendía a mantenerlas unidas. A veces conseguía separar las hojas sólo en una pequeña zona en una de las caras del cristal. Quizá se debiera ello a que entre las dos cortinas quedaba atrapada una capa de aire. Podían formarse así grandes y hermosas distorsiones de las cortinas de agua, con suaves curvas hacia arriba y hacia abajo. Sustituí el cristal del reloj por un tenedor y encontré distorsiones similares de múltiples cortinas, dondequiera que ellas se mezclaban.

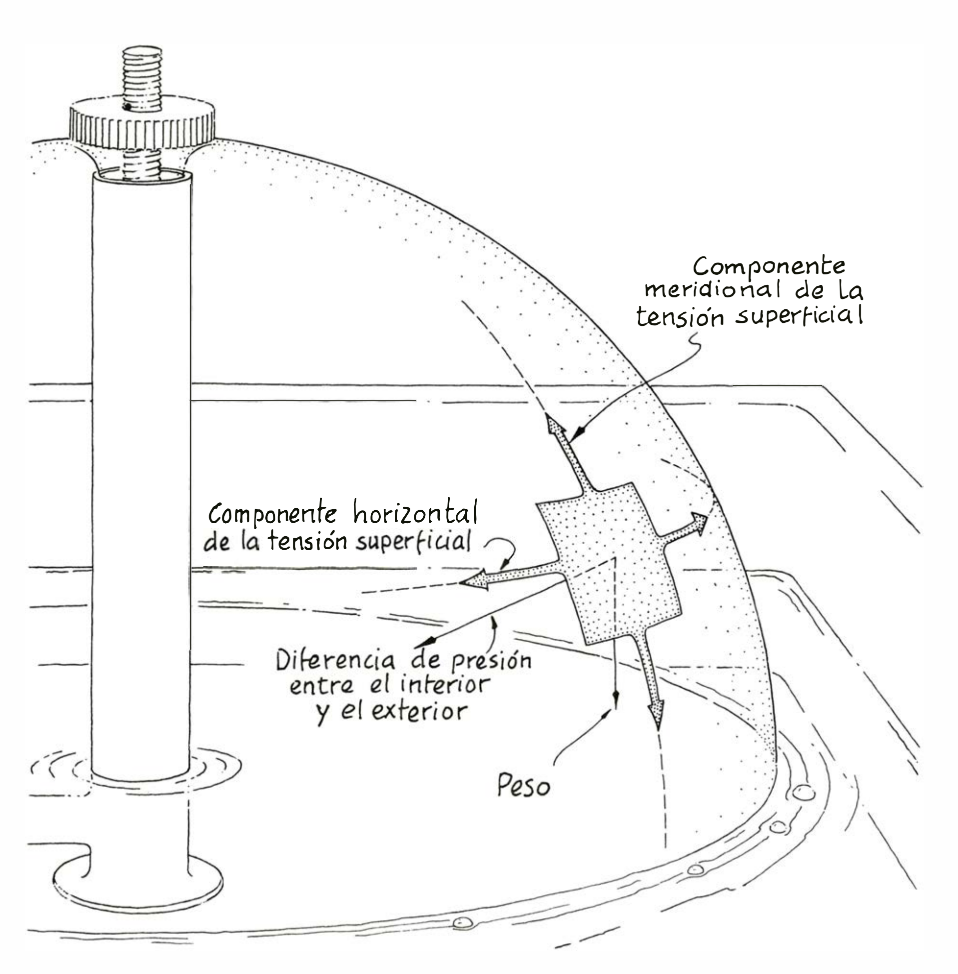
La forma de una campana está determinada por la diferencia de presión entre el interior y el exterior, por la gravedad, la viscosidad del aire en el interior de la campana y la tensión superficial. El efecto de la tensión superficial es doble: una fuerza hacia arriba y hacia abajo a lo largo del meridiano de la campana y una fuerza horizontal a lo largo de la circunferencia. Si la campana fuera esférica, las dos tensiones serían iguales, pero los diferentes radios de curvatura, según la dirección horizontal y la dirección meridional, hacen que dichos esfuerzos sean diferentes.

Si el lector investiga las cortinas y las campanas de agua esporádica o sistemáticamente, puede estar seguro que se divertirá bastante con ellas. También puede buscar nuevos moldes con los que generar figuras o bien crear esculturas de agua con múltiples campanas mezcladas unas con otras. Hay incluso otro camino para obtener cortinas de agua curvadas consistente en hacer caer agua por un borde como si fuera una cascada. También puede obtenerse una figura parecida a una campana vertiendo agua por las paredes de un cilindro vertical, de manera que la cortina que cae por la parte inferior del cilindro se una en un punto por debajo del mismo.

Si el lector fotografía las campanas de agua hechas en su fregadero, hallará que la cámara descubre muchas más turbulencias de las que se ven a simple vista. El único camino para eliminar este in-



Cambios en la forma de una campana de agua al incrementar la presión interior



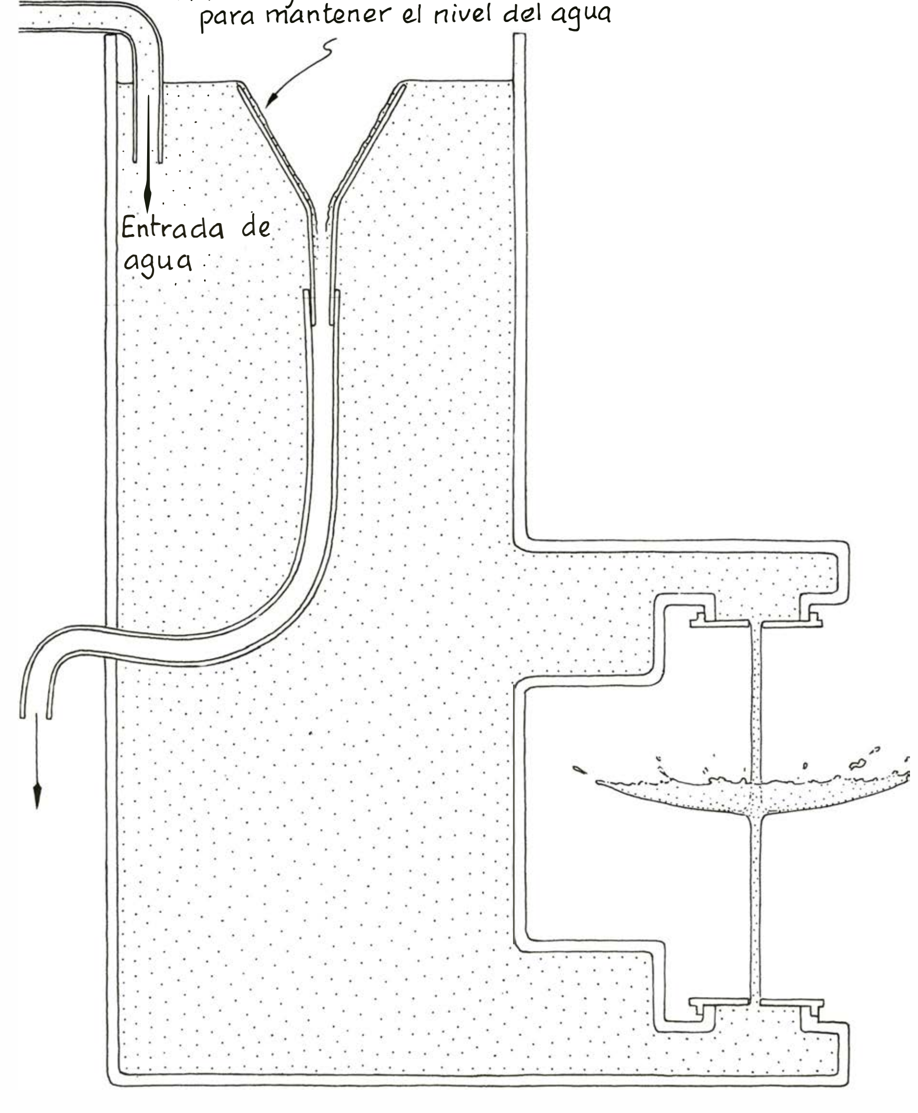
Fuerzas que actúan sobre una zona de la superficie de una campana de agua

conveniente es hacer pasar primero el agua a través de un número de delgados y estrechos tubos y también por varias hojas de cedales de alambre.

Las cortinas y campanas de agua pueden obtenerse también haciendo chocar dos chorros laminares de agua, bien de frente o bien inclinados. Bajo condiciones correctas, el agua se esparce formando una cortina que puede ser o bien plana (fuera del radio de desintegración) o bien curva. Ajustando el flujo y el diámetro de cada una de las corrientes, puede forzarse la curvatura en una determinada dirección: la cortina se curva hacia fuera de la corriente que lleva más cantidad de movimiento. Las cortinas casi planas aparecen cuando las cantidades de movimiento de ambas corrientes son las mismas. A una determinada diferencia entre los momentos, la cortina se pliega hacia atrás sobre uno de los chorros y origina una campana.

Se ha venido empleando para la construcción de cortinas este método desde que Savart escribiera por vez primera acerca de las campanas de agua. Calculó que el diámetro de las cortinas o de las campanas es proporcional a la presión del agua en las corrientes y proporcional, asimismo, al cuadrado del diámetro de las dos aberturas por donde sale el chorro. Las mediciones experimentales mostraron que el radio de desintegración coincidía con el calculado por Félix Savart.

J. H. Lienhard y J. C. P. Huang realizaron amplios estudios sobre choques de corrientes de agua en los años sesenta. Utilizaron un gran recipiente de agua de unos cincuenta centímetros de hondo para producir las dos corrientes. La profundidad podía ajustarse por un mecanismo de rebose que mantenía constante el nivel del agua. Dos salidas en la pared del recipiente llevaban incorporados dos tubos perpendiculares a dicha pared. Con gran precisión, se taladró un agujero en cada uno de los tubos de manera que cuando se llenaba de agua el recipiente, salían unos chorros de los agujeros en direcciones encontradas. El montaje, estable, tenía la ventaja de poder controlar la velocidad de las corrientes muy fácilmente ajustando la altura del agua en el recipiente. Los agujeros se reemplazaban fácilmente por otros agujeros con otros diámetros.



Aparato para producir chorros de agua que colisionen

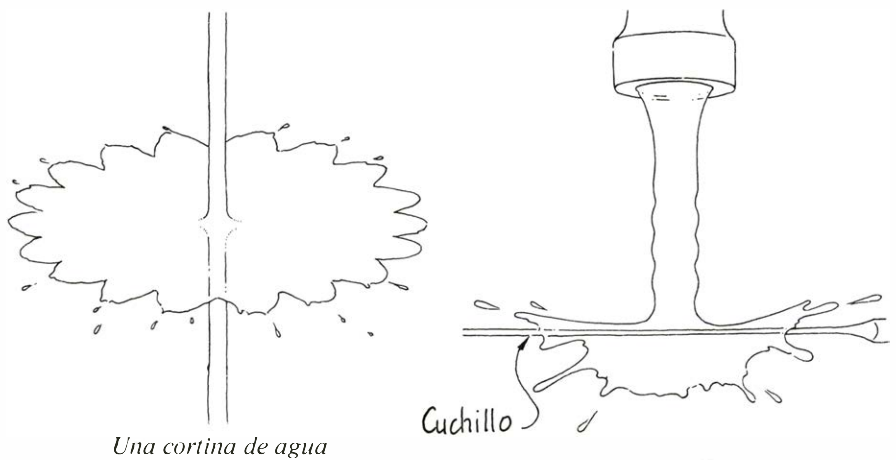
Las cortinas de agua pueden hacerse con menor dificultad, aunque se resiente el dominio del sistema. En éste, el agua viene desde el grifo a través de un distribuidor en *Y*, que enfrenta luego las dos corrientes, una en dirección a la otra. Si bien el montaje no permite modificar el diámetro del flujo, a no ser que cambiase las mangueras, sí puedo controlar la velocidad de la corriente por medio de las abrazaderas. Con este sistema, resulta casi un juego construir campanas de agua.

Por lo que se refiere a las cortinas, el rasgo que encierra mayor interés para los investigadores es su proceso de desintegración por los bordes. El trabajo más sobresaliente realizado sobre ese particular se debe a Lienhard y Huang, quienes experimentaron con cortinas formadas por dos corrientes en colisión. Algunos componentes de las desintegraciones sólo pueden verse en las fotografías tomadas a una gran velocidad o bien con películas, pero otros pueden verse en mis campanas de agua caseras.

Lienhard y Huang investigaron la desintegración en base a un factor conocido como el número de Weber, número adimensional que se aplica en los estudios de dinámica de fluidos. Dicho número se calcula multiplicando tres cantidades —la densidad, el diámetro de la corriente de fluido y el cuadrado de su velocidad— y dividiendo el resultado por la tensión superficial. En general, sólo pueden variar la velocidad y el diámetro. En las cortinas de agua de mi cocina, la velocidad puede regularse con la válvula del flujo; y, como todo chorro se va estrechando al caer, se puede elegir el diámetro adecuado en cualquier punto a lo largo de la corriente.

Con un número de Weber bajo (valores en un rango de 100 a 500), las cortinas son estables y tienen un perímetro casi circular, formándose gotitas de agua a lo largo del perímetro. En las fotografías hechas con alta velocidad se puede ver que estas gotitas se mueven a lo largo del perímetro coalesciendo unas con otras, aumentando su volumen, hasta que acaban por desprenderse constituyendo ya una sola gota. Las gotitas mayores y las gotas gruesas pueden observarse sin ayuda de ninguna fotografía, pero el movimiento de las gotitas a lo largo del perímetro es demasiado rápido para ser percibido.

Con números de Weber un poco más altos (entre 500 y 3000) el perímetro de

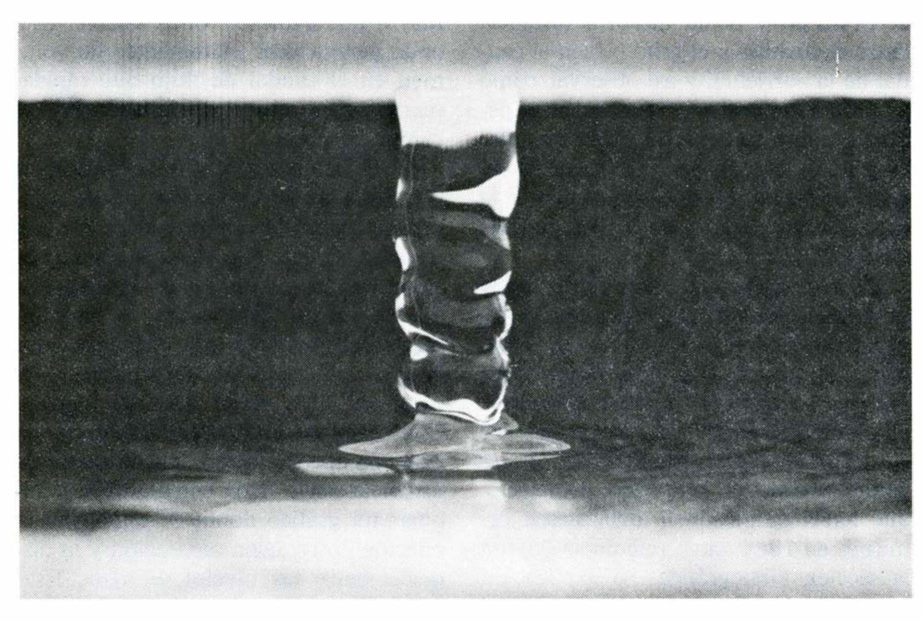


Una cortina de agua con un perímetro estrellado

Ondas en una corriente descendente

la cortina de agua adquiere una forma de estrella de muchas puntas, resultante de las ondas curvadas que se inician próximas al punto de colisión de las dos corrientes. Las ondas se propagan hacia fuera por la cortina dándole al perímetro una apariencia de brocado. Las cortinas con radio estable mayor se producen con un número de Weber entre 1000 y 2000. Próximo a 3000, los picos disminuyen y el perímetro se hace de nuevo circular. En tales condiciones puede verse cómo las ondas mayores se desplazan radialmente hacia fuera, desde el punto de colisión hasta el borde de la cortina. Estas ondas se llaman antisimétricas, debido a que el fluido en la cara opuesta de la cortina se mueve en dirección contraria, de suerte que en unos puntos de la onda la cortina pierde grosor y en otros gana. He podido mejorar la visibilidad de las ondas y de los picos del perímetro con números de Weber bajos, iluminando la cortina de agua con luz estroboscópica de alta frecuencia. Las ondas antisimétricas aumentan su amplitud con números de Weber mayores. La ruptura de los bordes puede observarse entonces sin dificultad porque la cortina empieza a flamear como si se tratara de una bandera que meciera el viento.

En mi montaje casero para conseguir corrientes que colisionen, cambiaba el ángulo de las corrientes al objeto de provocar colisiones más oblicuas. Cuanto menor era el ángulo bajo el que colisionaban las dos corrientes, tanto menor era la pieza de cortina que se formaba entre ambas y tanto mayor la porción que quedaba fuera del ángulo de las dos corrientes. (En la colisión, la mayor parte del agua sale despedida hacia dicha zona.) Cuando el ángulo entre las dos corrientes media, o se acercaba, a 60 grados, la cortina formada entre las dos corrientes era bastante menor que la que se formaba en la región exterior. La cor-



Ondas en una corriente descendente saliendo por un agujero abierto en un vaso de papel

tina mayor adquiría la figura de una hoja con la punta hacia fuera.

Volviendo a las cortinas de agua creadas por el choque de una corriente contra un obstáculo, he seguido de cerca las indicaciones de Sir Geoffrey Taylor (autor de varios trabajos sobre la dinámica de las cortinas de agua), quien introducía perturbaciones en el obstáculo. Tomé un cuchillo e hice un reborde radial en el fondo de un recipiente de plástico. Al servirme de éste para hacer cortinas de agua, con un número de Weber mediano, el reborde enviaba hacia fuera ondas curvas similares a las que se formaban, periféricamente, con un número de Weber similar.

Para proseguir los experimentos sobre cortinas formadas con obstáculos o con corrientes de colisión, convendrá profundizar en la correlación entre la velocidad del fluido y la formación de cortinas o bien su mecanismo de ruptura, necesitando para ello contar con un suministro de fluido cuya presión pueda controlarse, como hicieron Lienhard y Huang. También puede investigarse con otros fluidos distintos del agua. Algunos fluidos no newtonianos, como los que describimos en esta sección en enero de 1979, pueden resultar de particular interés, aunque deben evitarse los viscosos, ya que no se desplazarán según las exigencias requeridas para la formación de cortinas.

Hace unos siete años, la investigadora Elizabeth Wood (conocida por sus trabajos sobre los cristales y por su encantador libro *Sciencie for the Airplane Passanger*) me habló sobre otro fenómeno común de fregadero. Si mantenía el grifo abierto con un chorro tenue, de suerte que el agua fluyera lentamente, y con una hoja de cuchillo en la corriente cerca del grifo, podía ajustar la distancia entre el cuchillo y el grifo o bien el caudal de la corriente, y así observar cómo se producían ondas o rizos sobre el pequeño tramo de corriente existente entre el grifo y el cuchillo. La longitud de onda de los rizos u ondas dependía de la velocidad del agua y de la longitud del tramo de corriente.

Durante estos siete años, y de forma un tanto esporádica, he buscado una explicación para el fenómeno de estas ondas o rizos. La respuesta, finalmente, me vino sugerida por Lienhard mientras charlábamos sobre las campanas de agua para preparar este artículo. De repente recordó, a propósito del efecto, que había publicado la explicación del mismo en 1968, cuya referencia doy en la sección bibliográfica. La explicación contradice los trabajos hechos por Lord Rayleigh en 1878 y 1879 sobre la propagación de las ondas a lo largo de los cilindros de líquido. En la corriente descendente de agua pueden verse las ondas generadas por el impacto de la corriente con el cuchillo. Se mueven hacia arriba a lo largo de la corriente, con la velocidad a que desciende el agua de la corriente, lo que justifica la impresión de que las ondas no se muevan.

Para investigar esas ondas, Lienhard montó un recipiente lleno de agua que se vaciaba por un agujero central, en un vaso de martini lleno también de agua. El diámetro del agujero era de medio milímetro. La distancia de caída del agua era, a la vez, pequeña y constante, pero como el nivel de la misma en el recipiente disminuía lentamente, también la presión en el agujero y, por tanto, la velocidad del agua en la salida disminuían poco a poco.

Al decelerar la velocidad del agua de salida del recipiente, la correlación entre la velocidad de la corriente y la velocidad de las ondas indicaba que la longitud de onda de los rizos tenía que aumentar. La longitud de onda alcanzaba, por último, un límite superior. Rayleigh decía que la columna de agua sería inestable para cualquier onda con una longitud mayor que la circunferencia de la columna de agua. La mayor longitud de onda que puede darse en una onda estable en una corriente descendente es, pues, igual a la circunferencia de la columna. Cuando la velocidad del agua ha bajado tanto que resulta inestable, las ondas o rizos se rompen a veces en gotitas que salen disparadas sobrenadando la superficie del vaso de martini. La máxima inestabilidad se alcanza para una onda que tiene aproximadamente una longitud de onda igual a 4,5 veces el diámetro de la corriente. Constituye una longitud de onda más dilatada que la mayor que daba una onda estable. Esta onda genera tales incrementos de volumen en el caudal de flujo que rápidamente crece tanto que la corriente se rompe en gotas.

Si el lector dispone de una cámara con lente de aproximación puede fotografiar las ondas o rizos que se forman en una corriente que sale lentamente por el agujero de un recipiente. La longitud de onda de los rizos viene a ser inversamente proporcional a la distancia entre el nivel del agua en el recipiente y el nivel del agua en el vaso de martini. Con diversas fotografías hechas con la lente de aproximación, puede medir la distancia entre rizos consecutivos y dibujar un gráfico donde se establezca la relación entre dicha separación y la distancia entre los niveles de agua. Esto por lo menos es una buena manera de pasar menos aburrida una larga tarde de fin de semana.

Libros

Cuantificación en biología, la energía solar y el ecosistema de zonas áridas

Carlos M. Cuadras, Manuel de la Morena y Antonio Escarré

BIOMETRIA. PRINCIPIOS Y METODOS ESTADISTICOS EN LA INVESTIGACION BIOLOGICA, por Robert R. Sokal y F. James Rohlf, E. Blume, Madrid 1979. Uno de los objetivos de toda ciencia es poder llegar a explicar en términos cuantitativos los fenómenos que estudia. En 1623 Galileo afirmó que la naturaleza está escrita en lenguaje matemático. Y es bien cierto que el lenguaje y la metodología propia de las matemáticas constituyen la existencia misma de la física teórica y la mecánica racional. También la biología está influida por las matemáticas como lo está por la física y la química. ¿En qué medida? La matematización de una ciencia se alcanza cuando se rompe con el sistema tradicional de conseguir sus objetivos (descriptivo, empirista, etc.). Nace entonces una "nueva" ciencia que viene expresada, al menos en parte, en términos formalizados. Sin embargo, los intentos de matematización de la biología (Fisher, Haldane, Wright) sólo conducen a una biología matemática, que está en relación con la biología de la misma manera que la física matemática está en relación con la totalidad de la física, es decir, sin alcanzar a toda la ciencia biológica.

Se define la biometría como la ciencia que utiliza los métodos matemáticos, especialmente estadísticos, para analizar los fenómenos biológicos que se puedan expresar cuantitativamente. La contribución de la matemática a la biología es, por el momento, más fecunda y aceptada a través de la biometría que de la biología matemática. La biometría no propone unos principios y reglas fundamentales a partir de las cuales pueda derivarse una biología teórica. Se limita a utilizar la terminología y los resultados de la matemática y la estadística –que juegan el papel de instrumentos del método científico– y aplicarlos a la recogida, descripción y análisis de los datos experimentales, procedentes de la observación biológica. Los orígenes de la biometría están ligados a la antropología (tablas demográficas de J. Graunt y W. Petty, investigaciones antropométricas de L. A. J. Quetelet). El desarrollo más importante vendría de la mano de F. Galton, quien formularía los principios de la correlación y regresión para estudiar la herencia de los caracteres cuantitativos. Sus ideas serían expresadas de forma rigurosa por el genetista K. Pearson, creador de la estadística aplicada y promotor de una importante escuela biométrica, que tendría una enorme repercusión en el desarrollo de la estadística matemática. Su discípulo R. A. Fisher, también genetista, descubrió el análisis de la varianza y la mayoría de conceptos y métodos de la ciencia estadística actual, siendo además pionero del análisis multivariante.

En la actualidad, la biometría es una disciplina metodológica ampliamente aceptada por la mayor parte de los círculos científicos relacionados con la biología. Se puede afirmar, en líneas generales, que menosprecian las enormes posibilidades de los métodos biométricos solamente aquellos biólogos de perspectiva científica estrecha o demasiado tradicional, o por desconocimiento de la materia porque no estaba incluida la biometría en su formación.

R. R. Sokal y F. J. Rohlf ocupan un lugar de prestigio en el desarrollo de la biometría. Son artífices (junto con P. H. A. Sneath y C. D. Michener) de la taxonomía numérica, cuya influencia en la biología (sobre todo en ecología y microbiología) ha sido muy importante. *Biometría* es un interesante libro escrito por biólogos y dirigido a biólogos, que difiere de los textos normales de estadística. Hace un énfasis especial en las aplicaciones de la estadística a la biología, más que en la teoría, que queda reducida al mínimo. Para el biólogo, argumentan los autores, la utilidad de los métodos estadísticos es comparable al microscopio. Así como el conocimiento de la teoría no facilita el manejo del microscopio, conocer la estadística teórica no facilita la utilización práctica de los métodos estadísticos. Como sea que el libro es bastante completo, estando dirigido a un lector con poca base matemática, los autores hacen un notable uso de las "recetas". Tales "recetas" de estadística están resumidas en cuadros apropiados. Cada cuadro explica los pasos necesarios para realizar los cálculos que exigen un determinado método estadístico, a menudo sobre la base de unos datos numéricos. Para facilitar la búsqueda de la "receta" apropiada a un problema real, se incluye, en el apéndice del libro, una guía tabular de métodos estadísticos, con la misión de que el lector encuentre el cuadro apropiado, en función de las características de su problema. Sin embargo, la explicación complementaria a los cuadros de cada uno de los métodos es muy amplia, extendiéndose en la interpretación, las aplicaciones, y los problemas prácticos que puede presentar la utilización de un método a una situación concreta. El estilo descriptivo de las explicaciones está más cerca de un tratado de zoología que de un libro de matemáticas.

El primer capítulo es una hábil justificación de la necesidad de la biometría. Tomando como referencia la revista *The American Naturalist* (el editor es precisamente R. R. Sokal), los autores muestran cómo el porcentaje de trabajos publicados entre 1890 y 1960 con resultados no numéricos va disminuyendo, mientras aumentan los trabajos con resultados numéricos o acompañados de análisis estadístico. Argumentan entonces que los fenómenos biológicos solamente pueden ser estudiados dentro de la metodología y forma de pensar estadísticas. Los tres siguientes capítulos tratan de la forma y manejo de los datos en biología, seguida de una exposición de las principales nociones de estadística descriptiva. Para el manejo de los datos y realización de los cálculos, los autores suponen que el lector dispone de una calculadora de mesa o de un ordenador. Dedican un breve comentario a los ordenadores digitales y a las características de la programación en lenguaje FORTRAN. Incluyen asimismo un apéndice con 16 programas, que viene precedido por las instrucciones de utilización. Es este un enfoque muy positivo del libro, poco frecuente en obras del mismo nivel. La versión original (publicada en 1969) contenía además una serie de reglas prácticas para el manejo de calculadoras de mesa, manuales o electrónicas. La presente traducción de *Biometría* omite este tema, con la excusa de que el desarrollo de los ordenadores ha limitado el uso de las calculadoras de mesa. Esto es cierto respecto a las calculadoras antiguas. Sin embargo, a mediados de los años 70, han aparecido calculadoras programables de mesa o de bolsillo, con soporte de cassette o de tarjeta magnética, que funcionan de maravilla. La ma-

yoría de problemas de estadística univariante pueden ser procesados con estas máquinas, de utilización mucho más directa que un ordenador. Existen modelos cuyos fabricantes proporcionan excelente programotecas de estadística. Hubiera sido interesante que la traducción castellana planteara este tema debidamente actualizado, en lugar de optar por su eliminación.

Las nociones sobre la probabilidad son muy elementales. Se llega intuitivamente al concepto de probabilidad a partir del de frecuencia relativa. Las distribuciones binomial, Poisson y normal se introducen como leyes de probabilidad sobre los datos observados, eludiendo el concepto de variable aleatoria. Como los autores no definen el concepto de esperanza matemática, salvo una vaga referencia al valor esperado, las nociones de frecuencia esperada, parámetro poblacional, sesgo de un estimador, estimador insesgado, cuadrados medios, etc., se explican de una manera un tanto ambigua.

La parte más lograda de *Biometría* constituye los seis capítulos dedicados al análisis de la varianza. Los autores se encuentran en su mejor ambiente, dada la simplicidad y belleza del método, basado en modelos lineales, su formulación mediante sencillos cálculos algebraicos y las enormes aplicaciones que tiene en biología y agricultura. Distinguiendo entre los modelos de efectos fijos y los modelos de efectos aleatorios, desarrollan los diseños de un factor, dos factores y multifactoriales, los diseños anidados y una breve referencia a otros diseños. El procedimiento de cálculo asociado a cada método está explicado, en forma de "receta", en cuadros adecuados. Se encuentra a faltar el *S*-método de Scheffé en la sección dedicada a comparaciones múltiples. La referencia al análisis de la covarianza es demasiado corta. Los métodos para comprobar la verificación de las hipótesis básicas (normalidad, aleatoriedad, independencia de las réplicas, homoscedasticidad, etc.) necesarias para la correcta aplicación del análisis de la varianza, sirven para introducir al lector en la estadística no paramétrica. Las explicaciones teóricas son mínimas, pero la exposición de los métodos no paramétricos es suficiente para las aplicaciones.

Se mantiene también el esquema de los dos modelos (fijo y aleatorio) para explicar la regresión, según que la variable independiente sea controlable o aleatoria. Al exponer el tema de la correlación, se insiste en que es esencialmente diferente de la regresión. Tras explicar algunos coeficientes de correlación no paramétricos y sus pruebas de significación, la correlación se completa con una interesante referencia al análisis multivariante. Los autores ven en el análisis de datos el futuro más inmediato de la biometría.

El análisis de frecuencias presenta la particularidad de la utilización sistemática del estadístico *G*, basado en la razón de verosimilitud, además del estadístico *ji*-cuadrado. Los autores no disimulan su preferencia por el estadístico *G*, que relacionan con la cantidad de información de la muestra. Es sabido que ambos estadísticos son asintóticamente equivalentes. La comparación entre frecuencias observadas y esperadas en el caso de parámetros desconocidos se limita a los ajustes a las leyes binomial, normal y de Poisson, y a las pruebas de dependencia y homogeneidad en tablas de contingencia.

Biometría es una obra práctica dirigida a un lector poco exigente en conocer los fundamentos teóricos de la estadística. Sin embargo, recurrir sistemáticamente a procedimientos automáticos de cálculo, con insuficiente conocimiento de causa, puede degenerar en la aplicación incorrecta de los métodos propuestos, sobre todo cuando el material estadístico no se adapta del todo a ninguno. El desconocimiento de la teoría produce una incapacidad de maniobra, que puede conducir a aplicar el método "que más se le parece" al problema que debe resolverse. Es bien reconocida por la clase médica la peligrosidad de la "automedicación" por parte del paciente, si se nos permite hacer una comparación. De todos modos, por la notable extensión de los temas que trata, la obra constituye una buena introducción a la biometría y a la estadística aplicada. Es perfectamente recomendable, sobre todo si su estudio viene complementado con otra obra con mayor justificación teórica. Finalmente, es interesante mencionar que *Biometría* no contiene tablas estadísticas. Las tablas consultadas en el texto se encuentran en *Statistical Tables* (W. H. Freeman and Co., San Francisco, 1969), cuyos autores son también Sokal y Rohlf. (C. M. C.)

USO DIRECTO DE LA ENERGIA SOLAR, por Farrington Daniels; H. Blume, Madrid 1979. La utilización de la energía solar para fines industriales se remonta a un pasado ya lejano. La transformación de aguas saladas en dulces en el desierto de Chile en 1872, la construcción de diversos motores solares a finales del siglo pasado y principios de éste fueron experimentos, más o menos ingeniosos, precursores de una verdadera investigación que dio comienzo, puede decirse que seriamente, en la década de los cincuenta, especialmente impulsada por las investigaciones espaciales. El conocimiento de todo lo concerniente con el comienzo de esta investigación y el desarrollo de su primera mitad es evidente que resulta a veces imprescindible y siempre en extremo interesante para todos.

El doctor Farrington Daniels ha sido uno de los pioneros en los Estados Unidos en la investigación asociada a la transformación directa de la energía solar, y sus conocimientos y sus estudios fueron constatados en numerosos simposios internacionales celebrados sobre el tema por científicos de todo el mundo. El libro que comentamos sobre el uso directo de la energía solar es una resumida enumeración de las conclusiones y resultados, de las experiencias y de los conocimientos que adquirió el autor, en los muchos años dedicados a este estudio.

Tal vez sea esta la razón principal del interés que despierta la lectura del libro, al comprobar que se recibe información de primera mano, con todo el entusiasmo y la vívida experiencia de sus trabajos.

En el volumen se hace una descripción completa de todos los aspectos y distintas aplicaciones de conversión de la energía solar en mecánica, térmica y eléctrica, con la principal intención de señalar las muy diversas áreas de investigación que deberían seguirse en la actualidad y en el futuro. Repetidamente en todo el desarrollo del libro, y como una obsesión clara del autor, hay una invitación dirigida a todos los investigadores, para que con estas sugerencias puedan encontrar nuevas ideas y nuevas áreas de estudio.

La opinión del doctor Daniels sitúa a los países menos desarrollados económicamente como los que tienen una mayor urgencia de investigación sobre energía solar. Es evidente, de una parte, que los más industrializados también necesitan una investigación a más largo plazo, ya que sus recursos en combustibles, sus instalaciones de transporte de energía, etcétera, sólo hacen necesaria la sustitución, cuando aquellos combustibles comiencen a escasear, o cuando los precios sean prohibitivos. Por otro lado, los países menos desarrollados se encuentran en zonas donde la investigación debía ser ya rentable, pues los precios de los métodos convencionales son tan elevados que la energía solar resulta competitiva. Por último, como bien se señala en el libro, es en zonas rurales y soleadas, y sobre todo en zonas aisladas, donde la energía solar constituye una

fuente cargada de esperanzas y es también una gran ayuda la que los países industrializados pueden prestar a los que no lo están, al poder darles la posibilidad de ofrecerles investigación y desarrollo sobre las diferentes formas de utilización de la energía solar. Países pequeños, con poca experiencia en investigación y escasos recursos financieros, como opina Daniels, pueden, sin embargo, emprender una investigación tanto teórica como práctica en el uso de la energía solar.

Creemos que quien pretenda trabajar en este campo puede encontrar, en los distintos capítulos, la bibliografía básica que le permitirá situarse en el tema y conocer todo lo estudiado hasta 1964. Daniels reseña al final de cada capítulo la bibliografía completa y específica de las distintas posibilidades de investigación en los diferentes campos de aplicación. Los reseñados por el autor en cada capítulo se refieren a colectores de radiación solar, cocinas y hornos solares, calentamiento y destilación de aguas, calentamiento y enfriamiento de edificios, máquinas solares e incluso conversión termoeléctrica, termoiónica, fotovoltaica y fotoquímica entre los principales aspectos. La traducción española y su esmerada presentación se complementan con un anexo de datos útiles para España, sobre valores medios de radiación solar y un índice de entidades nacionales que trabajan en este campo.

Pero el libro también tiene gran interés, a nuestro juicio, para los que no estén directamente relacionados con el tema del uso directo de la radiación solar y sin embargo quieran informarse de manera veraz, sencilla y perfectamente entendible para cualquiera, de las posibilidades, de las esperanzas y del estado real de esa alternativa energética o "sistema prometedor", al que vale la pena dedicar todos los esfuerzos hoy comprometidos. Basta señalar que, conforme con las predicciones de Daniels, si en el año 1961 había sólo unas 20 viviendas solares en todo el mundo, en 1978 éstas pasaban de medio millar, y la producción de agua caliente que entonces se cifraba en el medio millón de instalaciones hoy puede sin duda llegar a los tres millones. En este sentido que apuntamos sobre la gran extensión de lectores a que puede interesarles el libro, hay una razón primordial que consiste en las breves y entendibles explicaciones para cualquiera que, al comienzo de cada capítulo, hace sobre los principios fundamentales de la aplicación considerada. Se trata de una concisa divulgación de tremenda altura científica y al mismo tiempo con la sencillez de exposición que el lector puede necesitar.

Por último, en la línea sumamente práctica de este volumen, el autor no olvida ir dando constantemente valores económicos actuales y futuros posibles, para orientar aún más en la línea más adecuada de investigación. Es evidente que un estudio de costes de un proyecto solar resulta difícil, mas para un especialista será de sumo interés comparar los costes dados por Daniels en este libro con los que en la actualidad rigen, en la siempre peligrosa extrapolación hacia el futuro (M. M.)

EL HOMBRE Y LA BIOLOGÍA DE LAS ZONAS ÁRIDAS, por J. L. Cloudsley–Thompson, traducción de la primera edición inglesa (1977) de Diorki traductores; 255 págs., 58 figuras (12 fotografías), Editorial Blume 1979. Cerca de la tercera parte de la superficie terrestre está constituida por zonas semidesérticas y desérticas, cuya producción primaria presenta cifras muy bajas, en muchos casos inferiores a las de ecosistemas acuáticos.

Cabe distinguir dos tipos de zonas desérticas: los desiertos polares, en los que la inasequibilidad del agua es debida a la congelación, y los desiertos propiamente dichos en los que existe un balance negativo de agua, al exceder la evaporación anual los valores de la precipitación. Es a este último tipo de desiertos a los que va dedicado el libro de Cloudsley-Thompson, excelente conocedor de buena parte de esas tierras, y que ha destacado como investigador de diferentes aspectos de la biología de los desiertos. El autor declara, como objetivo primordial de este libro, contribuir a la clarificación del problema que constituyen los desiertos, no sólo desde el punto de vista científico sino también desde el sociológico y económico. Una visión integral de las zonas áridas es probablemente el aspecto más destacable de este interesante libro, que consta de trece capítulos y un epílogo.

El capítulo primero se inicia con una tipología de los desiertos (subtropicales, costeros fríos, causados por barreras naturales y continentales interiores) y analiza los aspectos climáticos que originan el modelo de distribución de las zonas áridas. En el capítulo dos se tratan las particularidades edáficas de las zonas áridas y se dedica especial atención a las dunas que pueden alcanzar hasta 200 metros de altura, y cuyos complejos cubren superficies de muchos miles de kilómetros cuadrados. El capítulo tres subraya la heterogeneidad estratigráfica que constituyen en conjunto las zonas desérticas en contraste con una mayor uniformidad geomorfológica, y hace un esbozo de paleoclimatología de las zonas áridas.

La desertificación antropogénica es el tema tratado por el capítulo cuarto. El incendio sistemático del bosque, en la búsqueda de obtener una sabana más propicia para la explotación agrícola y ganadera, es una práctica que debió comenzar hace unos 50.000 años. Se subraya el hecho de que son muy pocos los pueblos primitivos existentes capaces de vivir en equilibrio en bosques tropicales, ya que la mayoría habitan en poblados y cultivan claros del bosque, lo que necesariamente requiere la previa tala e incendio. El empobrecimiento agrícola de ciertas zonas llevaría a la introducción del pastoreo, cuyo exceso ha contribuido de forma importante en la expansión de las zonas desérticas.

En el capítulo quinto, el profesor Cloudsley-Thompson hace una recopilación de pruebas fósiles e históricas para demostrar que el desierto del Sahara, tal como lo conocemos, es de reciente aparición. Sin perder de vista las variaciones climáticas, como causa principal de que se haya podido producir la extensión de este desierto, se insiste en la influencia humana en el desplazamiento de los cinturones de vegetación.

La biología de las zonas áridas aparece en los capítulos seis al doce. El primero de ellos trata de la vegetación deserticola y del conjunto de adaptaciones que presentan las plantas en medios de aridez extrema, e incluye los más importantes usos que los pobladores de estos territorios hacen de los materiales vegetales. Especialmente se pone de relieve la extraordinaria importancia que en estas zonas tiene el mantenimiento de una cierta cobertura vegetal, sin la cual la erosión del suelo se incrementa de forma exponencial.

El síndrome adaptativo de los animales de zonas desérticas, la termorregulación y las especiales características de los grandes mamíferos susceptibles de explotación constituyen la temática de los capítulos siete al nueve respectivamente. Se resaltan las singulares convergencias adaptativas entre animales y vegetales frente al estrés hídrico, y también entre distintos grupos animales que presentan sistemas muy similares para la regulación de las entradas y salidas de agua, y para el mantenimiento de un balance térmico adecuado. No se omite el clásico problema de la coloración negra especialmente frecuente en animales del desierto cuya significación parece ser primordialmente advertidora. Los mamíferos de gran tamaño, que no pueden recurrir para regular su temperatura a la ocultación en refugios, presentan una

serie de adaptaciones fisiológicas interesantes, como gran resistencia a la deshidratación, intervalo de variación de temperatura corporal más amplio y escasa producción de excretas. En relación con su explotación, se señala el escaso impacto sobre la vegetación que producen los animales salvajes de caza, muy móviles y siempre con bajos niveles de densidad y, por el contrario, el efecto nocivo de la utilización por pobladores sedentarios de la cabra, cuya explotación podría sustituirse por la más adecuada del camello.

Los rasgos fundamentales que permiten a la eurioica especie humana vivir en zonas de extrema sequía, aparecen en el capítulo diez; el amplio espectro trófico de los pueblos que habitan los parajes desérticos, es una consecuencia más de la escasa producción de materia viva. El capítulo once trata el problema de las plagas y de su extensión como consecuencia del intento de explotación de zonas semiáridas, de lo que son un claro exponente las plagas de langosta y otras especies de ortópteros, que presentan fases solitarias o gregarias diferenciables morfológicamente, dependiendo del espacio vital con que se encuentran las formas juveniles.

En el penúltimo capítulo se pasa revista a los principales organismos que pueden resultar dañinos al hombre, parásitos vegetales y animales y especialmente artrópodos y reptiles venenosos. Tienen especial importancia los problemas sanitarios derivados de la extensión de sistemas de irrigación.

En el capítulo final y el epílogo, el autor hace una serie de consideraciones sobre las posibilidades de explotación de las zonas áridas, de forma compatible con la conservación de la naturaleza. El nomadismo es considerado un sistema poco eficaz de explotación pero que si se favoreciese, en vez de facilitar el paso a la vida sedentaria, permitiría la explotación de pastoreo que de otra forma se perdería. Una ayuda al nomadismo que reuniese más información sobre las irregulares precipitaciones en esas zonas, e hiciese más predecibles las buenas rutas migratorias, podría compensar la tendencia a la formación de poblados sedentarios mucho más negativos desde el punto de vista ecológico.

En su conjunto, el libro constituye una buena síntesis de los conocimientos existentes sobre las zonas desérticas, sobre todo desde el punto de vista ecológico, pero sin ignorar otros aspectos del medio físico, así como históricos, sociológicos y económicos, y se completa con una selección bibliográfica adecuada. (A.E.)

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes;

TESTS BACTERIANOS DE SUBSTANCIAS POTENCIALMENTE CANCERIGAS

METHODS FOR DETECTING CARCINOGENS AND MUTAGENS WITH THE SALMONELLA/MAMMALIAN-MICROSOME MUTAGENICITY TEST. Bruce N. Ames, Joyce McCann y Edith Yamasaki en *Mutation Research,* vol. 31, n.º 6, págs. 347-364; diciembre, 1975.

PROPHAGE λ INDUCTION IN *ESCHERICHIA COLI* K12 *ENVA UVRB:* A HIGHLY SENSITIVE TEST FOR POTENTIAL CARCINOGENS. Patrice Moreau, Adriana Bailone y Raymond Devoret en *Proceedings of the National Academy of Sciences of the United States of America,* vol. 73, n.º 10, págs. 3700-3704; octubre, 1976.

ULTRAVIOLET MUTAGENESIS AND INDUCTIBLE DNA REPAIR IN *ESCHERICHIA COLI.* Evelyn M. Witkin en *Bacteriological Reviews,* vol. 40, n.º 4, págs. 869-907; diciembre, 1976.

PROGRESOS HACIA UN REACTOR DE FUSION TOKAMAK

TOKAMAK DEVICES. L. A. Artsimovich en *Nuclear Fusion,* vol. 12, n.º 2, págs. 215-252; marzo, 1972.

TOKAMAK RESEARCH. H. P. Furth en *Nuclear Fusion,* vol. 15, n.º 3, págs. 487-534; junio, 1975.

A PHYSICIST'S ABC ON PLASMA. L. A. Artsimovich. Mir Publishers, 1978.

RECENT PROGRESS IN TOKAMAK EXPERIMENTS. Masanori Murakami y Harold P. Eubank en *Physics Today,* vol. 32, n.º 5, págs. 25-32; mayo, 1979.

UNA ALDEA DE COLONOS PRENEOLITICOS A ORILLAS DEL EUFRATES

PAPERS IN ECONOMIC PREHISTORY. Dirigido por E. S. Higgs. Cambridge University Press, 1972.

FARMING IN PREHISTORY. Barbara Bender. St. Martin's Press, 1975.

THE EXCAVATION OF TELL ABU HUREYRA IN SYRIA: A PRELIMINARY REPORT. A. M. T. Moore en *Proceedings of the Prehistoric Society,* vol. 41, págs. 50-77; diciembre, 1975.

THE NEOLITHIC OF THE NEAR EAST. James Mellaart. Charles Scribner's Sons, Inc., 1976.

CORALES Y ARRECIFES CORALINOS

PROBLEMS OF GROWTH AND CALCIUM DEPOSITION IN REEF CORALS. T. Goreau en *Endeavor,* vol. 20, n.º 77, págs. 32-39; enero, 1961.

THE BIOLOGY OF CORAL REEFS. C. M. Yonge en *Advances in Marine Biology:* vol. 1, dirigido por F. S. Russel. Academic Press, 1963.

REEF CORALS: AUTOTROPHS OR HETEROTROPHS? Thomas F. Goreau, Nora I. Goreau y C. M. Yonge en *The Biological Bulletin,* vol. 141, n.º 2, págs. 247-260; octubre, 1971.

CORAL REEFS AND MOLLUSCS. C. M. Yonge en *The Royal Society of Edinburgh Transactions,* vol. 69, n.º 7, págs. 147-166; 1974.

SUPERNOVAS Y FORMACION DE ESTRELLAS

DARK NEBULAE, GLOBULES, AND PROTOSTARS. Bart J. Bok en *Publications of the Astronomical Society of the Pacific,* vol. 89, n.º 531, págs. 597-611; octubre, 1977.

OBSERVATIONAL EVIDENCE FOR SUPERNOVA-INDUCED STAR FORMATION: CANIS MAJOR R1. William Herbst y George E. Assousa en *The Astrophysical Journal,* vol. 217, n.º 2, parte 1, págs. 473-487; 15 de octubre de 1977.

PROTOSTARS AND PLANETS. Dirigido por T. Gehrels. University of Arizona Press, 1978.

PROBLEMAS FISICOS CON MUCHAS ESCALAS DE LONGITUD

CRITICAL PHENOMENA IN 3.99 DIMENSIONS. K. G. Wilson en *Physica,* vol. 73, n.º 1; págs. 119-128; abril, 1974.

THE RENORMALIZATION GROUP IN THE THEORY OF CRITICAL BEHAVIOR. Michael E. Fisher en *Reviews of Modern Physics,* vol. 46, n.º 4, págs. 597-616; octubre, 1974.

RENORMALIZATION GROUP: CRITICAL PHENOMENA AND THE KONDO PROBLEM. Kenneth G. Wilson en *Reviews of Modern Physics,* vol. 47, n.º 4, págs. 773-840; octubre, 1975.

INTRODUCTION TO THE RENORMALIZATION GROUP AND TO CRITICAL PHENOMENA. Pierre Pfeuty y Gérard Toulouse. John Wiley & Sons, 1977.

EL CONTROL HORMONAL DEL COMPORTAMIENTO EN UN LAGARTO

PSYCHOBIOLOGY OF REPTILIAN REPRODUCTION. David Crews en *Science,* vol. 189, n.º 4208, págs. 1059-1065; 26 de septiembre de 1975.

BIOLOGICAL DETERMINANTS OF SEXUAL BEHAVIOR. Dirigido por John B. Hutchinson. John Wiley & Sons, Inc., 1977.

THE PARENTAL BEHAVIOR OF RING DOVES. Rae Silver en *American Scientist,* vol. 66, n.º 2, págs. 209-215; marzo-abril, 1978.

ENDOCRINE CONTROL OF REPTILIAN REPRODUCTIVE BEHAVIOR. D. Crews en *Endocrine Control of Sexual Behavior.* Dirigido por Carlos Beyer. Raven Press, 1979.

ATMOSFERA Y CONTAMINACION ATMOSFERICA

AIR POLLUTION. R. Scorer. Pergamon Press, Oxford, 1972

AIR POLLUTION. H. C. Perkins. McGraw-Hill Book Co., New York, 1974.

CLIMA Y TIEMPO. H. Flohn. Biblioteca para el Hombre Actual, n.º 37. Labor, Barcelona, 1977.

CONTAMINACION ATMOSFERICA: FUNDAMENTOS FISICOS Y QUIMICOS. J. H. Seinfeld. Instituto de Estudios de Administración Local, Madrid, 1978.

JUEGOS MATEMATICOS

CONVERGENCE ON THE ARGAND DIAGRAM. L. W. H. Hull en *The Mathematical Gazette,* vol. 43, n.º 35, págs. 205-207; octubre, 1959.

COMPLEX NUMBERS IN GEOMETRY. I. M. Yaglom. Academic Press, 1968.

THE HISTORICAL DEVELOPMENT OF COMPLEX NUMBERS. D. R. Green en *The Mathematical Gazette,* vol. 60, n.º 412 págs. 99-107; junio, 1976.

TALLER Y LABORATORIO

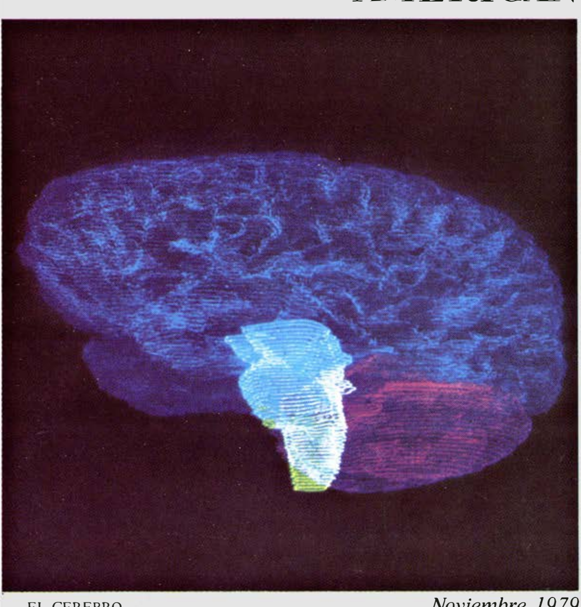
WATER BELLS. F. L. Hopwood en *The Proceedings of the Physical Society,* sección B, vol. 65, parte 1, n.º 385B, págs. 2-5; 1 de enero de 1952.

CAPILLARY ACTION IN SMALL JETS IMPINGING ON LIQUID SURFACES. John H. Lienhard en *Transactions of the American Society y of Mechanical Engineers, Series D. Journal of Basic Engineering,* vol. 90, págs. 137-138.

THE BREAK-UP OF AXISYMMETRIC LIQUID SHEETS. J. C. P. Huang en *Journal of Fluid Mechanics,* vol. 43, págs. 305-319; 28 de agosto de 1970.

EL PROXIMO NUMERO, EXTRAORDINARIO Y MONOGRAFICO, ESTARA DEDICADO AL CEREBRO

¿COMO FUNCIONA EL CEREBRO? ESTA SIGUE SIENDO UNA DE LAS MAYORES INCOGNITAS DE LA FRONTERA DEL CONOCIMIENTO. PERO YA PUEDEN APUNTARSE ALGUNAS RESPUESTAS GRACIAS A LAS INVESTIGACIONES LLEVADAS A CABO EN MUCHAS DISCIPLINAS CONVERGENTES, QUE DEMUESTRAN QUE ESTE ORGANO TIENE UNAS POSIBILIDADES MUCHO MAYORES DE LAS QUE SE LE SUELEN ATRIBUIR.



MECANISMOS CEREBRALES DEL MOVIMIENTO, Edward V. Evarts
El control de los movimientos del cuerpo requiere señales del cerebro a los músculos y viceversa.

EL CEREBRO, David H. Hubel
Introducción general a la neurobiología y a su problema fundamental: ¿cómo funciona el cerebro humano?

LA NEURONA, Charles F. Stevens
La célula nerviosa transmite señales a través del axón y las recibe en las dentritas.

MICROSISTEMAS NEURONALES, Eric R. Kandel
Dichos sistemas, estudiados en organismos simples, son capaces de formas simples de aprendizaje y memoria.

ORGANIZACION DEL CEREBRO, Walle J. H. Nauta y Michael Feirtag
¿Cómo surgen los departamentos cerebrales?

QUIMICA DEL CEREBRO, Leslie L. Iversen
En la transmisión de las señales nerviosas intervienen unas 30 micromoléculas.

MECANISMOS CEREBRALES DE LA VISION, David H. Hubel y Torsten N. Wiesel
Las investigaciones sobre el córtex visual primario descubren cómo el cerebro procesa la información sensorial.

DESARROLLO DEL CEREBRO, W. Maxwell Cowan
Asentamiento y disposición de las neuronas embrionarias.

RITMOS ELECTRICOS DEL CEREBRO, Elio García Austt y W. Buño, Jr.
El cerebro genera ritmos y se sirve de algunos de ellos para organizar la actividad motora y el estímulo sensorial.

ESPECIALIZACIONES DEL CEREBRO HUMANO, Norman Geschwind
Habilidades humanas como la del lenguaje requieren unas disposiciones del sistema cerebral de las que carece el resto de reino animal.

TRASTORNOS CEREBRALES, Seymour S. Kety
Algunos se deben a defectos congénitos o a traumatismos y, otros, a la interacción de factores genéticos o ambientales.

REFLEXIONES EN TORNO AL CEREBRO, F. H. C. Crick
Para entender el funcionamiento del cerebro habrá que cambiar nuestros métodos de análisis y reflexión en torno al mismo.